SPATIAL ANALYSIS OF REPRODUCTIVE HEALTH IN SYDNEY, NS

SPATIAL ANALYSIS OF SELECTED REPRODUCTIVE HEALTH OUTCOMES OF WOMEN LIVING IN THE VICINITY OF THE SYDNEY TAR PONDS, SYDNEY, NS

By

PATRICK F. DELUCA B.Sc. (Hons)

A Thesis

Submitted to the School of Graduate Studies

In Partial Fulfillment of the Requirements

for the Degree

Master of Arts

McMaster University

MASTER OF ARTS (2004) McMaster University
(School of Geography and Geology) Hamilton, Ontario

TITLE: Spatial Analysis of Selected Reproductive Health Outcomes of Women Living in the Vicinity of the Sydney Tar Ponds, Sydney, NS

AUTHOR: Patrick F. DeLuca, B.Sc. (Hons) (McMaster University)

SUPERVISORS: Dr. Pavlos S. Kanaroglou and Dr. Susan J. Elliott

NUMBER OF PAGES: ix, 173

ABSTRACT

Decades of steel production and coking in the community of Sydney, Nova Scotia, have led to severe environmental insult. Increased amounts of air pollution from Sydney Steel Corporation and toxic emissions from the Muggah Creek Watershed have been well documented in several studies of the area since the 1960s. This research examines the potential impacts of exposure to hazardous waste on the reproductive health of women living in Cape Breton Regional Municipality through the following objectives: i) to assess the spatial pattern of various types of adverse reproductive events, plausibly linked to the environmental exposure of interest; ii) to determine if this pattern is related to proximity to the Tar Ponds/Coke Ovens site. To address the first objective, point pattern analysis was applied to observations from the Atlee Perinatal Database to determine if the observed pattern exhibited any clustering. To address the second objective, a multinomial logistic regression model was employed to determine if proximity to the Tar Ponds/Coke Ovens site was an important covariate of the adverse outcomes under study (preterm births, low birthweights, congenital anomalies and stillbirths). The results of the bivariate K-function indicated that there was weak global clustering for preterm births for two different time periods, while the ratio kernel estimates demonstrated that the patterns of the outcomes were non-random even after correcting for the underlying population distribution. The results of the multinomial logistic model demonstrated that variables pertaining to maternal characteristics, pregnancy history, current pregnancy maternal diagnoses, neonatal measures were important explanatory variables in the analysis. Place of residence was an important

explanatory variable for preterm births and congenital anomalies. However, due to various limitations these results must be interpreted with caution.

ACKNOWLEDGEMENTS

This thesis would have been impossible to complete without the contributions of several individuals. First, I would like to thank my supervisors Pavlos Kanaroglou and Susan Elliott. Their guidance, support, patience (this did take four years to complete, albeit on a part-time basis) and friendship are very much appreciated. They have both provided me several opportunities, not only within the context of this research, but also with several other research projects and they have both left an indelible mark on me as a researcher. Thank you both.

I would also like to thank Henry Muggah, the principal investigator of the project, for his advice on reproductive health, and his suggestions not only for this thesis, but for other aspects of the project as well. I have learned a great deal from him and it has truly been a pleasure to work with him. I would also like to thank the other members of the research team: Bruce Wainman and John Eyles at McMaster University and Helen Mersereau and Gerrard MacKinnon at UCCB.

I would also like to thank Tara Burra and Bethany Haalboom. By working with them, I was able to gain more insights into the project and it helped in framing my research. A large debt of gratitude is also owed to Becky Attenborough and John Fahey at Nova Scotia Reproductive Care. Without them, I would not have any data for this thesis. I particularly thank John for his friendly assistance. We spent the better part of two days locked in his office extracting and cleaning data for 22108 pregnancies in CBRM.

Thanks to several members of the Centre for Spatial Analysis: Antonio Paez and Darren Scott for helpful suggestions with the categorical analysis; Jamie Spinney, Hanna Maoh and Dmitris Potoglu who acted as my sounding board for various ideas; Deane Maynard and Laura Labate for keeping things organized; and Ron Buliung for his years of friendship, and advice on both a personal and professional level.

I also appreciate the efforts of the other members of the research staff: Amber White, Theresa Odo, April Eby, and Anita Toth. The departmental staff has been tremendous as well, thank you to Ann Wallace for keeping me organized, Darlene Watson, Josie D'Allesandro, Luce Lavigne, and Janice Wade. Although not directly related to my thesis, without their efforts, other aspects of my professional life would have been more difficult.

Thanks to my family, for providing me support through all my endeavours and putting up with me during some of the more stressful periods of my life. Thanks to the members of the Hamilton Stars, for providing a fantastic outlet for my frustrations. Finally thanks to all my friends especially Sam, Ned, Jim, Jamie, Jerry, and Rick. You boys are the best friends anyone can ask for. You are all like brothers to me.

TABLE OF CONTENTS

Abstract....iii
Acknowledgements....v
List of Tables....viii
List of Figures....ix

CHAPTER ONE: INTRODUCTION....1
1.1 The Research Problem....1
1.2 Research Context....2
1.3 Research Objectives....2
1.4 Contributions of this Research....3
1.5 Chapter Outline....5

CHAPTER TWO: LITERATURE REVIEW....7
2.1 Introduction....7
2.2 Health and the Environment....8
2.3 Biological Pathways to Adverse Reproductive Health....14
2.4 Selection of Health Outcomes....19
2.4.1 Low Birthweights and Environmental Exposure....21
2.4.2 Preterm Births and Environmental Exposure....23
2.4.3 Congenital Anomalies and Environmental Exposure....24
2.4.4 Stillbirths and Environmental Exposure....26
2.5 Spatial Analysis and Health....27
2.6 Summary....30

CHAPTER THREE: METHODS....33
3.1 Introduction....33
3.2 Study Area....34
3.3 Sources of Data....38
3.3.1 Spatial Data....38
3.3.2 Reproductive Health Data....39
3.3.3 Census Data....42
3.4 Geocoding of the Atlee Perinatal Database....43
3.5 Analytic Methods....46
3.5.1 Exploratory Spatial Data Analysis....47
3.5.1.1 Visualizing Negative Reproductive Events....48
3.5.1.2 Bivariate K-Functions....49
3.5.1.3 Ratio of Kernel Estimates....53
3.5.2 Categorical Data Analysis....57
3.5.2.1 Descriptive Analyses of Adverse Reproductive Outcomes....58
3.5.2.2 Multinomial Logistic Regression Models....59
3.6 Summary....62

CHAPTER FOUR: RESULTS OF EXPLORATORY SPATIAL DATA ANALYSIS 63
4.1 Introduction 63
4.2 Geocoding and Visualization of Adverse Reproductive Events 64
4.2.1 Geocoding 64
4.2.2 Dot Maps 65
4.3 Exploration and Modeling of Spatial Point Patterns 69
4.3.1 Bivariate K-Functions 70
4.3.2 Ratio of Kernel Estimates 74
4.4 Discussion and Summary 81

CHAPTER FIVE: MULTIVARIATE ANALYSIS OF ADVERSE REPRODUCTIVE OUTCOMES 92
5.1 Introduction 92
5.2 Descriptive Analyses of Adverse Reproductive Outcomes 92
5.3 Results of Bivariate Analyses 95
5.4 Results of Multivariate Analyses 103
5.5 Discussion 108
5.6 Summary 121

CHAPTER SIX: CONCLUSION 122
6.1 Introduction 122
6.2 Summary of Findings 122
6.3 Research Contributions 129
6.4 Future Research Directions 132

BIBLIOGRAPHY 135

APPENDIX I: List of Commonly Used Abbreviations 148

APPENDIX II: List of Variables Obtained From the Atlee Perinatal Database 149

APPENDIX III: Outcome of Pregnancy on Last Follow-up as Coded in the Atlee Perinatal Database 158

APPENDIX IV: Bivariate K-Functions 160

APPENDIX V: List of Variables Used in the Analysis and Their Definitions 166

APPENDIX VI: Results of Categorical Data Analysis 170

LIST OF TABLES

Table 3.1 Types of geocoding used in the Atlee Perinatal Database..... 44
Table 4.1 Results of the geocoding process..... 65
Table 4.2 Results of the bivariate K-function for all outcomes..... 84
Table 4.3 Prevalence of pregnancy outcomes in Nova Scotia and Canada (after Burra, 2002)..... 88
Table 4.4 Prevalence rates per 1000 of reproductive outcomes as calculated from the Atlee Perinatal Database..... 88
Table 5.1 Odds ratios derived from Contingency analyses of adverse reproductive outcomes and proximity to the Coke Ovens site 96
Table 5.2 Comparison of mean distances of cases and non-cases to the coke ovens site..... 97
Table 5.3 Odds ratios derived from contingency analyses of adverse reproductive outcomes and various covariates (n=15800)..... 99
Table 5.4 Odds ratios derived from contingency analyses of adverse reproductive outcomes and various covariates among primiparous women (n=6009). 102
Table 5.5 Statistically significant CORs from the MNL model (n=15800)..... 104
Table 5.6 Statistically significant CORs from the MNL model for primiparous women (n=6009)..... 107
Table 5.7 Number of pregnancies contributed to the Atlee database by each woman (1988 – 2002)..... 116

LIST OF FIGURES

Figure 2.1 Potential exposure pathways in the vicinity of the Sydney Tar Ponds... 16
Figure 3.1 Location of Sydney, and other communities in Industrial Cape Breton. 35
Figure 3.2 Location of point sources of pollution in Sydney..... 36
Figure 4.1 Distribution of all births in CBRM, 1988-2002 (n=17648)..... 66
Figure 4.2 Distribution of preterm births in CBRM, 1988-2002..... 67
Figure 4.3 Distribution of congenital anomalies in CBRM, 1988-2002..... 68
Figure 4.4 Distribution of LBW in CBRM, 1988-2002..... 68
Figure 4.5 Distribution of stillbirths in CBRM, 1988-2002..... 69
Figure 4.6 Difference of K-functions for preterm births, 1988-2002..... 71
Figure 4.7 Difference of K-Functions for preterm births, 1988-1992..... 73
Figure 4.8 Difference of K-Functions for preterm births, 1993-1997..... 73
Figure 4.9 Difference of K-Functions for preterm births, 1998-2002..... 74
Figure 4.10 Ratio of kernel estimates for preterm births, 1988-2002..... 75
Figure 4.11 Statistically significant grid cells as determined through 99 simulations for preterm births..... 77
Figure 4.12 Ratio of kernel estimates for congenital anomalies, 1988-2002..... 78
Figure 4.13 Statistically significant grid cells as determined through 99 simulations for congenital anomalies..... 78
Figure 4.14 Ratio of kernel estimates for LBW, 1988-2002..... 80
Figure 4.15 Statistically significant grid cells as determined through 99 simulations for LBW..... 80
Figure 4.16 Statistically significant grid cells as determined through 99 simulations for stillbirths..... 81

CHAPTER ONE

INTRODUCTION

1.1 The Research Problem

Decades of steel production and coking in the community of Sydney, Nova Scotia, have led to severe environmental insult. Increased amounts of air pollution from Sydney Steel Corporation ((SYSCO), see Appendix I for all acronyms used throughout this thesis)) and toxic emissions from the Muggah Creek Watershed (commonly referred to as the Sydney Tar Ponds) have been well documented in several studies of the area since the 1960s (as cited in Guernsey *et al.*, 2000). The closure of the Coke Ovens has led to a decrease in air pollution; however, toxic materials remain in the Tar Ponds (as cited in Guernsey *et al.*, 2000). With over 700 000 tonnes of sediments contaminated with polyaromatic hydrocarbons (PAH's), polychlorinated biphenyls (PCB's), heavy metals, and volatile organochlorines (VOC's) in the watershed, the Sydney Tar Ponds/Coke Ovens site is considered one of the most contaminated sites in Canada (Environment Canada, 1999).

The role of the physical environment as a determinant of health is a major concern reported by both the public and policy makers (Elliott *et al.*, 2001). There have been a few studies to date that suggest that the environment may have had a health impact in Sydney and the surrounding area, in particular with respect to cancer (Mao *et al.*, 1985; Band and Camus, 1998; Veugelers *et al.*, 1999a and 1999b; Geurnsey *et al.*, 2000). It is also known that human reproductive health is intrinsically related to the environment, and is sensitive to adverse conditions (Mersereau *et al.*, 1999; DeWals, 1999). Adverse

reproductive outcomes result from one or several defects occurring in a complex developmental process that is determined genetically and influenced by the environment (DeWals, 1999). This research will attempt to explore the link between the contamination in the vicinity of the Tar Ponds and reproductive health of women in Sydney, Nova Scotia and the surrounding area.

1.2 Research Context

This thesis is one component of a larger programme designed to explore the impacts of the contamination in the Muggah Creek Watershed on reproductive and psychosocial health of Sydney residents. The research team is trans-disciplinary and employs mixed methods to address the following four components:

i. Investigating the incidence of abnormal early-pregnancy outcomes;
ii. Assessing exposure to environmental contaminants as indicated by body burden;
iii. Investigating residents' risk perceptions of and psycho-social responses to environmental threats to reproductive (and other) health; and
iv. Using spatio-temporal analysis within a Geographic Information System (GIS) to assess interactions and influences of the resultant factors on individual and community health.

This work falls under the umbrella of the fourth part of the research programme. Using data from a Provincial population-based database, this thesis describes the analysis and results of reproductive outcomes for women who reside in Cape Breton Regional Municipality (CBRM).

1.3 Research Objectives

Specifically, the goal of this research is to determine if the reproductive health of women in the Sydney area has been compromised due to environmental exposure from

living in the vicinity of the Sydney Tar Ponds. This goal will be addressed through the following objectives:

i. To assess the spatial pattern of various types of adverse reproductive events, plausibly linked to the environmental exposure of interest;
ii. To determine if this pattern is related to proximity to the Tar Ponds/Coke Ovens site.

Meeting these objectives will contribute to the existing body of literature surrounding the relationship between environmental exposures and reproductive health. Initially, this research was to be presented to the Health Studies Working Group (HWSG) of the Joint Action Group (JAG) in Sydney. This committee comprised of local residents, business people, politicians and youth was formed in 1996 with the goal of remediating the Watershed (JAG, 2003). However, this group has since been disbanded. The results of this thesis will still be made available to the individuals that have subsequently taken their place. Through dissemination of the results in this manner, the research team will help Sydney residents to develop an understanding of the possible relationship between the environment in the watershed and reproductive health within their community.

1.4 Contributions Of This Research

In most studies revolving around industrial sources of pollution and reproductive health, researchers have found mixed results. Some studies point to an association between environment and adverse reproductive outcomes while others do not. Methodologically, this research will contribute to the existing body of literature in several ways. First, the use of spatial analysis to determine if adverse outcomes tend to cluster in the Sydney area, specifically around the Tar Ponds/Coke Ovens site, is unique to the area.

Studies of this nature have never been conducted in Sydney and the results of this phase of the analysis will help in gaining insight to the problem and add to the body of literature surrounding health effects of the area. In addition methodological improvements will be made to some of the spatial analytic tools available, particularly to the methods proposed by Reader (2001). A second contribution lies in the type of modeling used in this research. The statistical modeling employed in this analysis is categorical in nature, specifically making use of multinomial logistic regression (MNL) models. While models of this type have been employed in studies dealing with discrete choices for the type of reproductive health care sought after (Glei and Goldman, 2000), it has never been applied in studies dealing with reproductive health outcomes and the environment, so this work will add to the current body of literature (both for reproductive studies, and for environment and health studies where the range of possible outcomes may be multinomial in nature).

Substantively, the analysis will help to elucidate the impact of industrial sources of pollution on the reproductive health of women in Sydney, and research of this type forms a component of JAG's site assessment process. This thesis describes the first attempt at employing spatial analysis to assess patterns of adverse reproductive outcomes in Sydney. In addition, it is one of the few studies that use spatial analysis as a tool to answer questions about reproductive health, rather than using reproductive health data to demonstrate the utility of a newly devised method. Traditionally, there have been two bodies of researchers investigating the use of spatial analysis in health. The first group is researchers with questions relating to health, and has a sense that GIS and spatial analysis

will help them in answering their questions, but may not have the capability to employ these methods effectively. The second group are primarily statisticians or methodologists who find health data useful to demonstrate a new method, but have little knowledge of health and health frameworks. This research attempts to bridge the gap between the two groups by being able to couple spatial analysis with informed decisions based on the environment and health literature.

1.5 Chapter Outline

The next chapter begins with a discussion of the theoretical context of this study. Included in this is a discussion on reproductive health and the environment with subsections devoted to possible biological pathways, and the selection of health outcomes for study. This is followed by a critical appraisal of past investigations of reproductive health impacts of community exposure to industrial pollution. The chapter concludes with a discussion on the use of GIS and spatial analysis in health and reproductive health research.

Chapter three discusses the methods applied in the thesis. The study area is described and includes key flashpoints in Sydney's industrial history. This is followed by a discussion on data sources and limitations of the Provincial database that is employed in the analysis. Following this, the various analytic tools used to assess point patterns arising from inhomogeneous populations are discussed. This is followed by the categorical data analysis employed in this research.

Chapters four and five present the research findings. The results of the point pattern analysis are presented in chapter four. Chapter five presents the findings of the

bivariate and multivariate analysis performed on the population database used to investigate the impact that distance from the Tar Ponds has on reproductive health in Sydney and the surrounding area.

In chapter six, the major findings are reviewed and followed by a discussion of the major contributions made by this research. The thesis is concluded with recommendations for future research.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

The continuously evolving interaction between humans and the environment has produced a mixture of beneficial and harmful products. While industrialization has produced a large variety of goods in mass quantities, it has also produced the largest number of environmental risk factors in the history of the human race. Indeed, humankind has introduced several elements into the environment that either pollute or modify environmental conditions with resulting negative impacts on diverse aspects of the ecosystem, including human health. In some instances, the environmental impacts are fairly constant over long periods of time (e.g. chronic exposure to air pollution), while in other cases impacts are rapid (e.g. environmental disasters). Bajaj *et al.,* (1993) illustrate how three environmental disasters may have adversely affected reproductive health. They discuss the release of methyl mercury into the Minimata Bay area in Kyushu, Japan in 1955, the release of methyl isocyanate as a result of the Union Carbide accident in Bhopal, India in 1984, and the contamination of cooking oil in Taiwan with PCBs in 1979. As a result of each of these accidents, increases in various negative reproductive outcomes were reported ranging from increases in spontaneous abortions to Congenital Minimata disease, which does not manifest itself until the infant is approximately 6 months of age (Bajaj *et al.,* 1993). While it may be possible to establish a link between health and the environment as a result of an industrial accident, it is much more complex to assess the impact of chronic exposure to toxic chemicals that are produced in everyday

industrial processes. A great deal of current environment and health research has centered on studies with weak associations such as the impact of air pollution on respiratory health, where small relative risks have been found to be statistically significant (e.g. Burnett *et al.*, 1995, Delfino *et al.,* 1997). It is in these studies where weak associations are examined that the limitations of the science involved are most pronounced.

Currently, there are over 60000 chemicals in commercial use, but toxicological information is only available for very few of these. For example the Agency for Toxic Substances and Disease Registry (ATSDR) has developed toxicological profiles on only 269 substances (ATSDR, 2003). Sullivan (1993) states that there are even fewer reproductive toxicological profiles completed. This chapter reviews the literature pertaining to the reproductive health impacts of chronic exposure to industrial pollution. In reviewing this literature, a brief discussion on environment and reproductive health will be covered, including definitions and possible biological pathways for adverse reproductive events to occur. This will be followed by a review of empirical studies across a range of exposures and reproductive outcomes. Also included in this chapter is a brief overview of the application of GIS and spatial analysis in health and reproductive health research.

2.2 Health and the Environment

Geography is the study of the earth as the home of humankind and as such it is primarily interested in the areal variation of physical and human phenomena on the surface of the earth (Fellmann *et al.,* 2001). A key characteristic of the discipline is the

study of the environment. Initially, research involving the environment was related to hazards. It has since broadened and as Negro-Vilar (1993) states:

> "Environment represents the totality of physical, chemical, biological, behavioural and socioeconomic factors or conditions that constitute the external milieu surrounding the human organism" (p. 59)

Health, defined by the World Health Organization (WHO), is "a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity" (WHO, 1948). This definition was expanded in 1986 and is now defined as:

> "...not only the absence of disease, but also the extent to which an individual or group is able, on the one hand, to realize aspirations and satisfy needs, and on the other, to change or cope with the environment" (WHO, 1986)

Gatrell (2002) adds that health can be thought of as "being physically and mentally fit and capable of functioning effectively for the good of the wider society" (see Elliott, 1999 for a review of the evolution of health).

The health of a population is multi-factorial in nature. An important factor is that of status. Evans (1994) makes the astute observation that "top people live longer" and are generally healthier. He illustrates that individuals who are economically well off, and higher on the social ladder tend to be healthier than those below them. In addition to this, there is a gradient observed in health through the social classes from the top to the bottom. A recent report from Health Canada confirmed this by providing evidence that as one goes up each "rung on the income ladder" Canadians have improved health, less sickness and longer life expectancies (Health Canada, 1999). This report went on to table several items in what it lists as the determinants of population health in Canada. Among these are: income and social status; social support networks; education and literacy;

employment and working conditions; social environments; physical environments; personal health practices and coping skills; healthy child development; biology and genetic endowment; health services; gender; and finally, culture (Health Canada, 1999). Each of these determinants is important on their own, but as Evans and Stoddart (1990) point out, they are all inter-related. As an example, Health Canada (1999) describes the trend seen with low birth weight (LBW) babies and maternal income. The effect occurs not just for the most disadvantaged group. As mothers move up the income scale, they have babies with higher birth weights on average than those below them. This means that problems are not just a result of poor maternal nutrition or poor health practices associated with poverty. Factors such as coping, empowerment, status and future orientation all come in to play.

The idea that health and geography (and environment) are linked is not new and as Gatrell (2002) points out, the two are inextricably linked. Whether it is accessibility to different health care options, or proximity to a point source of pollution, location is important. Medical geography is the branch of human geography that deals with the geographic aspects of health (status) and healthcare (systems). It seeks to improve our understanding of the various factors that affect the health of populations and individuals. A related discipline of medical geography, spatial epidemiology, is concerned with examining the distribution of disease and death at various geographic scales, in an attempt to determine if the presence or absence of a particular illness is associated with some factor(s) in the social or physical environment.

The combination of environment and health is also intimately related to medical geography. Agius (2001) defines environmental health as:

> "...comprising those aspects of human health, including quality of life, that are determined by physical, chemical, biological, social and psychosocial factors in the environment. It also refers to the theory and practice of assessing, correcting, controlling and preventing those factors in the environment that can potentially affect adversely the health of present and future generations."

This research is rooted in both spatial epidemiology and environmental health to examine reproductive health outcomes in the area of Sydney, Nova Scotia.

Michal *et al.* (1993) defines reproductive health as a "condition in which the reproductive process is accomplished in a state of complete physical, mental and social well-being". Similar to the definition of health, this entails more than absence of disease or disorders associated with the reproductive process. Savitz and Harlow (1991) point out that successful reproduction involves several components. These include:

> "the biologic processes of the male and female that yield viable sperm and a viable egg; the social and biologic interactions between the male and female that yield a viable conceptus; the biologic process within the mother and fetus that result in appropriate growth and development of the fetus; and the processes that ensure successful parturition" (p. 159)

The disruption of any of these processes can interfere with reproduction and consequently, multiple effects may be seen.

As mentioned previously, there is a great deal of evidence to suggest that one of the factors that may influence reproductive health is socio-economic status (SES) or social class. Evans (1994) discusses social status or hierarchies in the context of mortality, however the same ideas may apply to reproductive health. In fact, Vrijheid *et al.* (2000) explore this idea in examining the incidence of congenital anomalies. Using a

deprivation index, they found that the most deprived quintile of the index had a 40% higher risk of non-chromosomal congenital anomaly than the most affluent quintile after adjustment for confounding factors. In addition, they found that a gradient of increasing risk with increasing deprivation existed for neural tube defects (Vrijheid *et al.,* 2000). Stephansson *et al.* (2001) found a similar gradient when evaluating the risk in stillbirth in Sweden using occupational class (i.e., unskilled blue-collar workers, skilled blue-collar workers, low-level white collar workers and high-level white collar workers) as opposed to a deprivation index.

The idea that SES has an impact on reproductive health is not new (see Leon, 1991; Pattenden *et al.,* 1999 for example) as most studies attempt to control for it in some way (for example, Vrijheid et al., 2000). In fact, if one examines epidemiological studies on any of the adverse outcomes of pregnancy, a major factor that is common to many of them is the influence of parents' SES. Sullivan (1993) notes that poorer members of the population have the worst outcomes; however, SES alone cannot account for any adverse effect in pregnancy but may be a surrogate for other factors critical in affecting development. Michal et al (1993) suggest that SES is a surrogate for poor housing conditions, living in deprived areas (associated with poor hygiene, inadequate supply of good water, lower quality of indoor air, higher level of exposure to pollutants and infections), malnutrition, limited access to health care, social drug use (alcohol and smoking) and educational status of the mother.

In terms of other determinants associated with adverse reproduction, stress, or the in-ability to cope with it is an important factor. This is due to the fact that a prolonged

stress response can result in a variety of forms of physiological damage. In fact, stress has been found to induce infertility in males and early pregnancy failure in females (Negro-Vilar, 1993). Michal *et al.* (1993) point out that stress is one of the major contributors to adverse reproductive outcomes along with infection, malnutrition, chemicals and radiation.

When examining Sydney, Nova Scotia and the surrounding area of CBRM, many of the causal factors highlighted by Evans (1994) are at play. Rainham (2002) describes the area as "a culturally rich, but economically poor and government-dependent community". Once the steel factory ceased operations in 1999 (see Chapter 3 for more details on the study site and the industrial history of the area), all prospects of maintaining economic independence in the community were lost (Rainham, 2002). When comparing this area to the rest of the Province of Nova Scotia, it can be seen that the average unemployment is greater (22.5% in CBRM versus 13.3% in Nova Scotia), the dwelling values are lower ($63700 versus $86500) and the number of individuals with lower educational levels are greater (15.4% versus 12.6% of residents with less than grade 9 education) (Statistics Canada, 1996a).

Rainham (2002) points out that there is large body of evidence that links the steel industry with a variety of respiratory disorders, cancers and other impairments. Several studies have been conducted in the Sydney area examining various health outcomes with the most attention given to cancer (Mao *et al.*, 1985; Band and Camus, 1998; Veugelers and Guernsey, 1999a and 1999b; Geurnsey *et al*, 2000). Veuglers and Guernsey (1999a) also found increased rates of cardiovascular disease. It is also known that human

reproductive health is intrinsically related to the environment, and is sensitive to adverse conditions (Mersereau *et al.,* 1999; DeWals, 1999). Dodds and Seviour (2001) conducted a study examining the incidence of adverse reproductive outcomes in Sydney and the surrounding area. Their study suggests higher rates of congenital anomalies among Sydney residents than in the rest of Nova Scotia. The next section highlights the possible biological pathways through which this increase in adverse reproductive outcomes may be brought about.

2.3 Biological Pathways to Adverse Reproductive Health

Health Risk Assessment is the method used to determine the magnitude and probability of an adverse health effect from exposure to a substance. The first step in assessing the risk from toxic chemicals involves the selection of the contaminants of interest (Malachowski, 1995). Results from occupational and animal studies serve as the basis for identifying chemicals that are potential health hazards to humans. Based on epidemiological and animal studies, the risk assessments that have occurred involving the Tar Ponds in Sydney include petroleum hydrocarbons (benzene, toluene, xylene), PAHs, heavy metals (i.e., lead) and PCBs (JDAC Environment, 2002). Toxicity tests in experimental animals suggest that the aforementioned contaminants detected in the Tar Ponds are reproductive toxicants (Sullivan, 1993; Gomez and Mattison, 1995; Foster *et al.,* 1996). For example, Foster *et al.* (1996) describe several animal studies conducted by Health Canada to determine the reproductive toxicity of various congeners of PCBs, while Sharara *et al.* (1998) highlight that several studies show fetal loss, malformations and preterm births in rodents when exposed to PCBs. Birnbaum (1995) shows that

pregnant women exposed to dioxin-like compounds experience an increase in stillbirths, while Fein *et al.* (1984) and Rylander *et al.* (2000) have found links between exposure to PCBs and increased low birth weight. Heavy metals such as lead have been linked to intrauterine fetal death or preterm delivery (Sharara *et al.*, 1998; Lindbohm, 1992) and mercury has been linked to birth defects (Myers *et al.*, 1995)

Once the chemicals of interest have been established, the routes of exposure need to be identified. In Sydney, JDAC Environment (2002) have listed several categories of individuals that may have been exposed and the circumstances of their potential exposure. For the Coke Ovens in particular, they report that workers (industrial, construction and commercial) and off-site residents may have been exposed to chemicals via incidental ingestion of the soil, incidental dermal contact, and inhalation of vapours and dust. For the Tar Ponds, they add ingestion of fish as well as incidental contact with the water found in the Tar Ponds (JDAC, 2002). There are other potential routes of exposure that are presented in figure 2.1 (modified after Mackay and Webster, 1998).

This figure depicts the major contributors to the poorer environment experienced by the residents around the Tar Ponds and the typical routes that the pollutants would enter the body. An argument can be made for exposure through air, water and soil. The most direct route to the body is through the inhalation of pollutants. The main culprit historically would have been the Coke Ovens, however these ceased operations in 1988 (see Chapter 3 for more information). The fact that the main polluter is no longer operational is a limitation in this study; however, although no longer a direct contributor, the coking process has left a legacy in its wake through deposition of pollutants in both

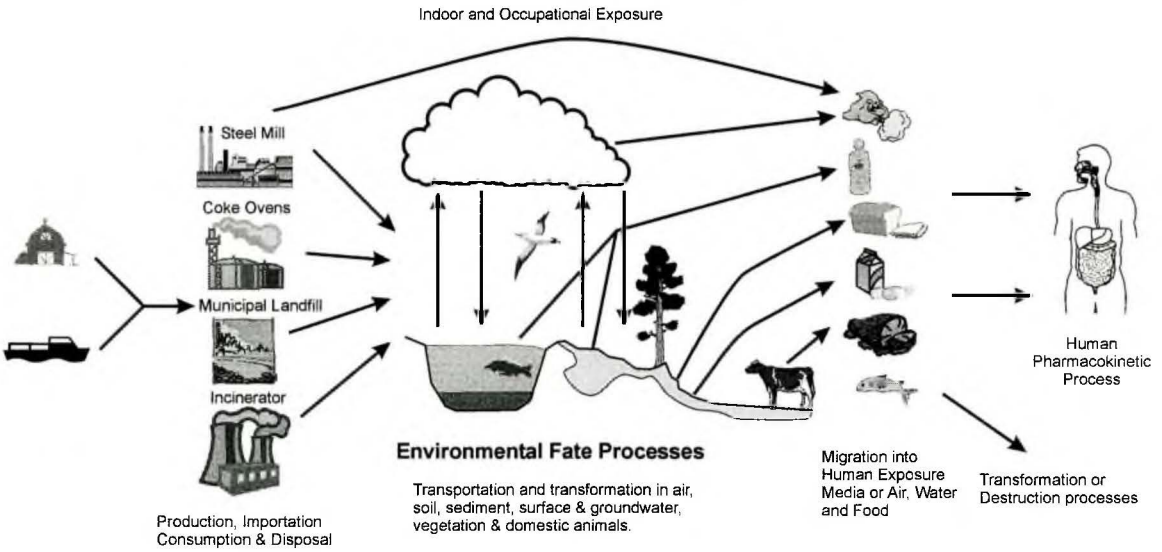


Figure 2.1. Potential exposure pathways in the vicinity of the Sydney Tar Ponds

soil and water. Additionally, De Wals (1999) notes that some agents may persist in the maternal body and produce an effect long after the exposure has occurred. Recently, the soils of some residents in the Whitney Pier neighbourhood in Sydney were tested for elevated arsenic levels by Health Canada (Nova Scotia Department of Health and Cape Breton District Health Authority, 2001)[1]. The other sources of air pollution include: the steel mill (operations ceased in 2001); the municipal waste site and the incinerator used to burn waste. Although not as significant as the coking process, the steel mill would have contributed to decreased air quality through emissions of particulate matter from blast furnaces. The municipal waste site, which contributes to the degradation of air quality through the release of biogas would also have an impact. In fact Goldberg *et al.* (1995)

[1] The results of these tests are unavailable to the public. They are accessible only by requesting them from the owners of the properties where the tests were conducted.

found elevated levels of low birth weights in residents exposed to biogas in Montreal, Quebec.

The other two routes of entry to the body are through dermal contact and the gastrointestinal tract. With respect to dermal contact, three possibilities exist, airborne deposition, direct contact with soil and direct contact with water. Of these, direct contact with soil and water are stronger than airborne deposition (Gochfeld, 1995). The gastrointestinal tract can also be a significant route of entry. The drinking of contaminated water, and the consumption of contaminated fish have been linked to increased adverse reproductive outcomes (Butler and Kalasinski, 1989, Colborn *et al.*, 1993). Although PCBs have been banned long ago, they are persistent in the environment and they bio-accumulate as they move up the food chain. PCBs are prevalent in the Tar Ponds and although the fisheries in the area were forced to close in the early 1990s, it is still conceivable that people consume seafood from the area today (it is unlikely however, that they drink the water). In addition to the consumption of seafood, fruits and vegetables from the garden may also provide a route of exposure if not properly washed before consumption. Although the consumption of fish is the best-documented route of exposure, there is considerable evidence that PCBs can become volatile in contaminated bodies of water such as the Tar Ponds (Chiarenzelli *et al.*, 2000) and wet sediments that surround it (Bushart *et al.*, 1998; Chiarenzelli *et al.*, 1996). These volatile PCBs can then be absorbed through inhalation

Once in the body, environmental exposures to chemicals follow the sequential pathways of pharmokinetics regarding absorption, transport, metabolic biotransformation,

tissue distribution, storage and excretion (Bajaj *et al.,* 1993). Savitz and Harlow (1991) suggest that every component of the reproductive process is susceptible to an exogenous agent. These agents can alter the reproductive system both directly and indirectly (Mattison, 1983) and since there is a multifaceted vulnerability to environmental insults Savitz and Harlow (1991) argue for the examination of multiple end points, as is the case in this study. The direct action of exogenous agents is through endocrine disruption. Direct effects occur if a chemical is structurally similar to an endogenous molecule and becomes capable of entering the reproductive system. Under this scenario, the chemical may alter normal cellular processes that may result in abnormal tissue growth, function or death (Savitz and Harlow, 1991). Exposures during the first trimester of pregnancy may lead to a miscarriage, or a congenital anomaly. Second or third trimester exposures might produce complications of pregnancy, fetal growth or timing of delivery (Savitz and Harlow, 1991). The endocrine system is also mentioned by Evans (1994), where he describes how it and the immune system can be modified by a prolonged stress response. The implication deduced by Evans (1994) is that individuals at the lower levels of the social hierarchy are often in more stressful situations and may have depressed immune and endocrine systems as a result. Once the immune system has been compromised, the individual becomes more susceptible to environmental insult. It is through the immune system that reproductive function may be altered indirectly (Mattison, 1983). The immune system reacts and responds to antigenic exposures and in turn can modulate or adversely affect reproductive events (Negro-Vilar, 1993). These indirect effects occur if a chemical requires metabolic conversion within the body before it is capable of exerting

a toxic effect. Some chemicals may mimic or block natural hormone action, thereby altering the reproductive process. Therefore, this indicates that toxins that directly affect the reproductive tissues themselves, but also substances that affect a number of other tissues, which indirectly regulate or support reproductive function, can impair reproductive health.

2.4 Selection of Health Outcomes

Toxicological studies usually involving animals would normally help to define the outcomes chosen for study. However, given that there are very few reproductive toxicological studies completed (Savitz and Harlow, 1991), epidemiological studies were used instead to inform the choice of outcomes examined. The existing literature demonstrates that human reproductive health is very sensitive to adverse environmental conditions (Savitz and Harlow, 1991, Bajaj *et al.,* 1993, Taskinen, 1993, Hertz-Piccioto *et al.,* 1996, Sharara *et al.*, 1998). Sullivan (1993) points out that exposures during the embryogenesis period in the first trimester may lead to miscarriages, congenital anomalies or stillbirths. The fetogenesis period (second and third trimester) might be expected to produce intrauterine growth retardation, preterm birth and complications with delivery (Sullivan, 1993). Finally, Savitz and Harlow (1991) point out that exposures in the third trimester may be related to low birth weight and preterm birth.

Based on the observations of Savitz and Harlow (1991) (i.e., every component of the reproductive process is susceptible to an exogenous agent and as such, that multiple reproductive end points should be studied), the outcomes selected for study in this thesis include; stillbirths, congenital anomalies, preterm births and low birth weights. In

addition, the outcomes selected were in part driven by the need to address the community concerns regarding human health impacts of the Tar Ponds. As such reproductive endpoints such as sperm quality, irregular menstrual cycles, infertility and time-to-pregnancy studies may be more difficult to quantify in a study such as this. In other words, the outcomes chosen for study have definitions (see Chapter 3 for the standard clinical definitions applied in the study) that make them easily observable and reliably recorded in the population based provincial database that stores information on every pregnancy in the province since 1988.

Prior to the broader research program, of which this thesis is a part, there was one other reproductive health study conducted in Sydney and the surrounding area (Dodds and Seviour (2001). This study yielded mixed results and used the aforementioned population based surveillance database as its source of data. Dodds and Seviour (2001) examined a range of reproductive health outcomes including congenital anomalies, low birth weight, preterm birth and intrauterine growth restriction. Although limited by the ecological study design, their results indicated statistically significant, elevated levels of congenital anomalies in Sydney versus neighbouring communities. The other outcomes, although in most respects higher in Sydney versus the surrounding areas, failed to yield statistically significant results. The authors concluded that since Sydney and the neighbouring areas shared a similar risk factor and socio-demographic profile, other factors likely explain the increased rates of congenital anomalies in Sydney. One of the factors that may set Sydney apart from the neighbouring communities in terms of risk is that of exposure to the Tar Ponds and the other point sources of pollution in the vicinity.

Burra (2002) examined some of the same outcomes (low birth weight, preterm births, stillbirths, congenital anomalies, miscarriage and infertility) in Sydney in a companion thesis. Using a survey, she conducted a retrospective cross-sectional study consisting of a sample of 500 women. Her results revealed a relationship between preterm births and proximity to the Tar Ponds using the straight-line distance from maternal residence to the Tar Ponds as a proxy for exposure.

The literature surrounding environmental exposure from hazardous industrial waste, landfills and incinerators and negative reproductive health in other areas has employed varying study designs, producing varying results. The majority of these studies involve one point source of pollution, usually in response to community concerns (similar to the case in Sydney). A brief summary of the literature for the outcomes under study is presented in the following section.

2.4.1. Low Birth Weight (LBW) and Environmental Exposure

LBW is defined by WHO as a birth weight less than 2500 grams and has been found to be associated with genetic, demographic, lifestyle and environmental factors (Kramer, 1987a; Kramer, 1987b; Kramer *et al.,* 1990). Ha *et al.* (2001) used logistic regression to study LBW in Seoul, South Korea and found an association with CO, NO2, SO2 and TSP. The relative risks, though small, proved to be significant when correcting for gestational age, maternal age, education, parity and infant sex. A major limitation of the study however is the lack of data on tobacco use and alcohol consumption. Additionally, there was not a clear point source of pollution so exposure was assigned on the basis of residential location within an administrative zone with an air monitor

measuring ambient air quality. Rogers *et al.* (2001), Gouveia *et al.* (2004), and Lin *et al.* (2001) also found positive associations with ambient air pollution and LBW. These studies were all similar in that exposure was based on monitoring sites in zones of residence. Perera *et al.* (2003) also found significant increases in LBW from ambient air pollution, however, their study employed personal monitoring devices to determine the level of PAHs that each pregnant woman was experiencing in New York City.

Dolk *et al.* (2000) on the other hand, had clearly defined point sources of pollution as they examined the impact of coke works in Great Britain on birth weight. They created zones of 7.5 km around 22 coking facilities and examined the outcomes within each zone. They conducted their modeling using logistic regression and used distance to the closest coking facility as the measure of exposure. However, they found no evidence of an increased risk of LBW with proximity to a facility. Berry and Bove (1997) on the other hand, found increased odds of LBW (odds ratio (OR) of 5.12 (2.14, 12.27 95% CI)) during periods of peak exposure to volatile organic compounds and heavy metals from living in proximity to a hazardous waste site in New Jersey. Goldman *et al.* (1985), also found significantly elevated ORs among exposed homeowners in the Love Canal area of Niagara Falls, New York. Baibergenova *et al.* (2003) tested the hypothesis that residents in a zip code with a PCB hazardous waste site or a water body contaminated with PCBs are at an increased risk of giving birth to a LBW baby. Again, location of residence was used as a proxy for exposure. Using logistic regression and controlling for maternal age, race, weight, height, education, income, marital status and smoking, there was still a statistically significant increased risk of giving birth to a male

LBW baby. Elliott *et al.* (2001) found small excess risks of LBW in populations living near landfills using a geographical comparison of individuals living within 2 kilometers of a landfill site versus those who do not. Finally, Goldberg *et al.* (1995) also found elevated rates of LBW around a landfill site in Montreal. Their study defined the exposed population in a manner similar to Baibergenova *et al.* (2003) using postal codes adjacent to the site.

While several studies point to elevated rates of LBW in areas near contaminated sites, there are just as many that fail to find any evidence of the impact of these sites on LBW. As mentioned previously, Dodds and Seviour (2001) failed to find statistically significant elevated rates between exposed (residing in Sydney) and unexposed (residing elsewhere). Bhopal *et al.* (1999) also failed to find any evidence that living close to major industries led to increases in LBW. Defining exposure by creating three zones (A, B and C with A being closest to industry and C being furthest) they actually found the opposite to what they expected (i.e., a reverse gradient existed, where the furthest area had the highest LBW and the closest had the fewest). Finally using landfills as the primary point source of pollution, both Fielder *et al.* (2000) and Baker *et al.* (1988) failed to find any evidence of increased LBW in exposed areas.

2.4.2 Preterm Births and Environmental Exposure

Kramer *et al.* (1998) describe preterm birth as arguably the most important maternal and child health problem in developed societies. WHO defines preterm birth as being born before 37 weeks gestational age or before 259 days (Moutquin, 2003). Three main conditions explain preterm birth: medically indicated preterm birth, preterm

premature rupture of membranes (each accounting for approximately 25% of all preterm births) and spontaneous preterm birth (accounting for approximately 50% of all preterm births). Currently recognized risk factors for each of these conditions include medical conditions such as pregnancy-induced hypertension, unstable fetal condition, and maternal age (all risk factors for medically indicated preterm birth), infection and disadvantaged populations (preterm premature rupture of membranes) and various lifestyle factors (spontaneous preterm births) (Kramer, 2003; Moutquin, 2003, ACOG, 2001, Kramer *et al.*, 1998). Recently, the environment has been thought of as another potential risk factor. Similar to low birth weights, mixed results again appear when attempting to discern the impact of environment on preterm births.

Yang *et al.* (2002) found a significant excess of preterm deliveries (OR 1.18 (1.04-1.34 CI) for primiparous women using unconditional multiple logistic regression to model exposure to petrochemical and petroleum industries. Exposure was based on residence within three kilometers of a petrochemical complex. Although they found a significant effect with exposure, they were missing two key confounders, smoking habits and any meaningful values to assess the mother's socio-economic status. Berry and Bove (1997) also report an elevated OR in their retrospective follow-up study in New Jersey, while Goldberg *et al.* (1995), Goldman *et al.* (1985), Kharrazi *et al.* (1997) and Dodds and Seviour, (2001) found no elevated rates of preterm birth in their respective studies.

2.4.3. Congenital Anomalies and Environmental Exposure

Major congenital anomalies are defined as a structural abnormality with surgical, medical or cosmetic importance (Holmes, 1999). Vrijheid *et al.* (2002a) have found a

33% increase in the risk of congenital anomalies among residents near hazardous waste sites in a European collaborative study. As mentioned previously, Dodds and Seviour (2001) found statistically significant elevated rates of congenital anomalies in Sydney versus the surrounding communities. Dolk *et al.* (1998a) conducted a case-control study and demonstrated elevated ORs for women living within 3 kilometers of a hazardous waste site for several types of anomalies in Denmark, France, Belgium, Italy and the United Kingdom. They found that residence within 3 kilometers of a landfill site was associated with a significantly raised risk of congenital anomaly controlling for maternal age and socioeconomic status (OR 1.33 (95% CI 1.11-1.59)). They also noted a consistent decrease in risk with distance away from the sites. In a related study involving chromosomal congenital anomalies, Vrijheid *et al.* (2002b) found an OR of 1.41 (1.00-1.99) using the same methods as Dolk *et al.* (1998a). Fielder *et al.* (2000) and Elliott *et al.* (2001) found small excess risks of congenital anomalies in populations living near landfill sites in Great Britain. Finally, Dummer *et al.* (2003), using a retrospective cohort study found an increased risk of lethal congenital anomaly around incinerators and crematoriums in northwest England.

Bhopal *et al.* (1999) using a geographical comparison design found no evidence of increased risk from living in close proximity to industrial sources of pollution in England. Similar to Dolk *et al.* (1998a), this study used residential proximity to industry to proxy for exposure. Sosniak *et al.* (1994), Marshall *et al.* (1993) and Croen *et al.* (1997) also found no association between exposed versus unexposed to hazardous waste

sites and congenital anomalies using case-control studies and maternal residence as the exposure measure.

2.4.4 Stillbirths and Environmental Exposure

A stillbirth occurs when the product of conception is extracted from the mother without any signs of life. To be considered a stillbirth, the duration of the pregnancy is at least 20 weeks of gestation, prior to that and the non-viable fetus is considered to be a miscarriage (Kramer *et al.*, 2002). About one fourth of all stillbirths have unknown causes and a few studies indicate that pollution may play a role. Several studies relating to stillbirths have shown little or no elevated levels. The few that have shown an effect include Dummer *et al.* (2003), and Landgren (1996). Dummer *et al.* (2003) found an increased risk of stillbirths (OR 1.04, 95% CI: 1.01-1.07) for women residing near crematoriums in Cumbria, England. Landgren (1996) found that the odds ratio of stillbirths was twice as high in a municipality with the highest arsenic exposure in a county in southern Sweden.

Dolk *et al.* (2000) find no evidence of elevated stillbirths in an area with 32 coke works in England. Bhopal *et al.* (1999) and Sosniak *et al.* (1994) also found no evidence to support a hypothesis that living close to industry may lead to increased levels of stillbirth.

From this review it can be concluded that there is some evidence from the literature to suggest a possible association between exposure to industrial waste and adverse reproductive outcomes. However, as most authors of these papers concluded, it

is difficult to determine if there is a causal relationship between the exposure and an increase in adverse reproductive outcomes.

2.5 Spatial Analysis and Health

GISs are being used with increasing frequency to analyze the spatial distribution of health outcomes (Forand *et al.,* 2002). As mapping and spatial analysis software become easier to use, public health professionals are increasingly employing these tools to aid in cluster investigations and disease surveillance. Several authors highlight the utility of GIS and spatial analysis within the health field (see Gatrell and Senior, 1999; Higgs and Gould, 2001; Rushton, 2003 for example) and Dunn *et al.* (2001) point out that GIS and spatial statistics tend to complement classic epidemiological studies. With respect to the application of GIS and spatial analysis to reproductive health outcomes, the literature is sparse. There are several spatial studies that use adverse reproductive health data, however, the majority of them are authored by methodologists who find utility in the use of reproductive health data to demonstrate a method. For example, Forand *et al.* (2002) use data on congenital anomalies from New York State to examine data quality issues. They could have just as easily used hospital separations or mortality due to any condition, however, as previously noted by Dolk (1999) reproductive health data are continually updated. Rushton and Lolonis (1996) used birth defect data in Iowa to demonstrate a clustering technique involving simulations. Building on this, Reader (2001) used data on low birth weights in Florida to demonstrate a method to find statistically significant clusters of high rates of low birth weights using kernel density

estimation. As a final example, Kammann and Wand (2003) demonstrate the utility of geostatistics when dealing with non-linear relationships between response variables and predictors using low birth weight as an example.

There are comparatively fewer studies that make use of spatial analysis to address a research question involving adverse reproductive outcomes. Dolk *et al.* (1998b) and English *et al.* (2003) are two examples where spatial analysis is employed to address questions of clustering. Dolk (1999) advocates the use of spatial analysis both for the analysis of disease clusters as well as for disease surveillance. She recognizes that studies dealing with reproductive outcomes have several advantages over spatial studies involving chronic diseases such as cancer or heart disease. First, and foremost, births are finely geo-referenced facilitating disaggregate analysis, a significant improvement over the aggregate analysis in ecological frameworks. Gatrell *et al.* (1996), among others, demonstrate several techniques that can be used to analyze the spatial properties of data of this type. The second advantage that Dolk (1999) highlights is that births are continually updated in databases, therefore providing a proper denominator when calculating rates of a disease. Finally, when exposure is residential proximity to a pollutant source, exposure misclassification due to migration is minimized, since the period between exposure and diagnosis is short (Dolk, 1999). This does not however, correct the problem for chronic exposure and body burden, which may produce an effect long after the exposure occurred (De Wals, 1999).

De Wals (1999) highlights the fact that it is difficult to pinpoint a single environmental cause when analyzing adverse reproductive outcomes, and states that only

five cluster studies to date have been attributed to a single factor. This suggests that the majority of clusters are either artifacts or are produced by the accidental congruence of different and unrelated causes. However, De Wals (1999) does not deny that environment has an impact, and it is the aim of this work to determine how much of an impact, if any, it has on reproductive health in Sydney.

A major limitation of all of the mentioned studies in section 2.4, including Dodds and Seviour (1999) is that they fail to capture the importance of spatial variation. Spatial variability is important in studies of this nature for various reasons. First, it is essential to understand how the mean (first order effect) varies through space. There will be instances where the data may exhibit clusters of events, and conversely, pockets where no events are occurring, possibly signaling under-reporting of events or some other sort of bias (Forand *et al.,* 2002). As populations, behaviours, genes and adverse environmental conditions are not uniformly distributed over space, clusters of adverse reproductive outcomes can be expected. Second, it is important to understand how the underlying correlation structure varies through space (second order effect). Since spatial behaviour is usually a mix of both first and second order effects, it is important to examine each in this study.

Different forms of spatial analysis are used in this thesis to address the first two research questions posed in the previous chapter. In so doing, spatial analysis is used for substantive rather than methodological purposes, thereby further contributing to that body of literature. In addition, as Dunn *et al.* (2001) point out, spatial analysis is a useful complement to epidemiological studies and as such, helps in addressing the third research

question. Finally, reproductive outcomes in Sydney have been examined previously, but never from a spatial perspective, so this research will add to that body of literature surrounding the impact of the Tar Ponds on human health, helping to fulfill one of the mandates of the Health Studies Working Group of JAG.

2.6 Summary

While the impacts of environmental exposure from an industrial accident on reproductive health have been demonstrated, the results of epidemiological studies dealing with low-level exposures over a long period of time have been equivocal at best. This chapter began with an examination of reproductive health and the environment and provided the necessary background to establish a link between the environment and the potential impacts on reproductive health. Most authors agree that environment does have a role to play in the etiology of adverse reproductive outcomes, although the exact mechanism of action is still up for debate. This was followed by a section detailing the biological pathways of the pollutants in the Sydney area. It is conceivable that pollutants may enter the body through air, ingestion and dermal contact. The primary polluter, the Coke Ovens, is no longer in operation and this forms a limitation since exposure from inhalation is probably the most direct route that a contaminant has to enter the body (Gochfeld, 1995). Taking that into consideration, the amount of contaminants remaining in and around the Tar Ponds is well documented and may in fact pose a risk through ingestion, and dermal contact with soils not only on site, but soils in resident's yards that may have traces of contaminants through deposition. Inhalation may still play a role through the volatilization of PCBs and evaporation of surface waters from the Tar Ponds.

In this thesis, exposure is estimated by proximity to the contaminated site. Although this is only a proxy for exposure, and a poor one at that, Dolk (1999) highlights that it is the most frequently applied method to assess exposure and under the circumstances is the only method available to assess exposure to local residents.

The third section of this chapter reviewed the substantive findings from past studies of reproductive health impacts of exposure to industrial waste. Studies employing various designs, varying exposure measures and reproductive health outcomes, produced equivocal results. From this review it can be concluded that there is some evidence from the literature to suggest a possible association between exposure to industrial waste and adverse reproductive outcomes. In this thesis, the relationship between environment and reproductive health will be explored in a multinomial framework, using different categories of birth outcomes as the dependent variable. All of the studies reviewed here examined only one health outcome at a time, with the dependent variable represented in a binomial form. A review of the literature failed to produce any studies that presented various reproductive health outcomes in a multinomial framework, so this research will help to fill in some gaps in the literature.

The final section described the use of spatial analysis in reproductive health research. To date, reproductive health data has largely been used by methodologists to illustrate the effectiveness of new methods in spatial analysis, but rarely has it been used to address a substantive research question. This thesis will use spatial analytical techniques to address three research questions involving the impact of the environment on reproductive health in Sydney. This spatial analysis will complement the existing

studies conducted in the larger research project. In addition, a spatial study has never been conducted in Sydney; so this is an innovation in that regard, while at the same time contribute to the literature surrounding the use of spatial analysis with reproductive health.

The remaining chapters of this thesis describe a spatial study of reproductive health in Sydney, Nova Scotia. The next chapter presents the methods employed in this thesis.

CHAPTER THREE

METHODS

3.1 Introduction

This chapter describes the methods utilized in this thesis to address the research objectives outlined below. As discussed in chapter one, these research objectives are a component of part four (iv) of the broader research programme; namely, to apply spatio-temporal analysis within a GIS to assess the interactions and influences on reproductive heath in Sydney and the surrounding area. This involves the statistical analysis of a population – based database to address the following research objectives:

i. To assess the spatial pattern of various types of adverse reproductive events, plausibly linked to the environmental exposure of interest;
ii. To determine if this pattern is related to proximity to the Tar Ponds/Coke Ovens site.

The next section of this chapter describes the study area and provides some context for the research. Burra (2002) provides a detailed community profile, and the aim of this section is not to reproduce those efforts, but rather to set the stage for this research by describing the industrial history and the point sources of pollution in the area. This is followed by a section discussing the sources of data used in this study, with particular attention given to the Atlee Perinatal Database and the steps taken to derive the outcomes and to geocode the observations. In the fifth section of the chapter, the statistical analyses applied in this thesis are described in detail. The section is sub-divided into two broad categories, one covering exploratory spatial data analysis (ESDA), and the other, categorical data analysis.

3.2 Study Area

CBRM is one of four counties located on Cape Breton Island in Nova Scotia, Canada. The majority of CBRM's residents are located in the north end of the county, where 6 of the 8 former municipalities can be found (i.e., Sydney, Glace Bay, New Waterford, Sydney Mines, North Sydney, and Dominion) (Figure 3.1). In 1995, these entities plus the remainder of Cape Breton County and Louisbourg were amalgamated into CBRM in order to consolidate resources, and to improve infrastructure and the provision of municipal services (Burra, 2002). Since amalgamation, CBRM is now the second most populous municipality in the province next to Halifax, with approximately 109 000 residents occupying a geographic area of approximately 2500 square kilometers (CBRM, 2003). The largest of the former municipalities is Sydney (approximate population of 26 000), which is located on the southwest arm of Sydney Harbour and is approximately central to the other communities that comprise what is known as Industrial Cape Breton (ICB) (Figure 3.1). Following Guernsey *et al.,* (2000), ICB is comprised of the major communities (listed previously) surrounding Sydney Harbour.

Industrial activity has a long history in Cape Breton, beginning with the French mining coal in the Sydney area as early as the 1700s (Burra, 2002). The steel industry, which has made Sydney famous, began with the construction of the steel plant in 1899 on

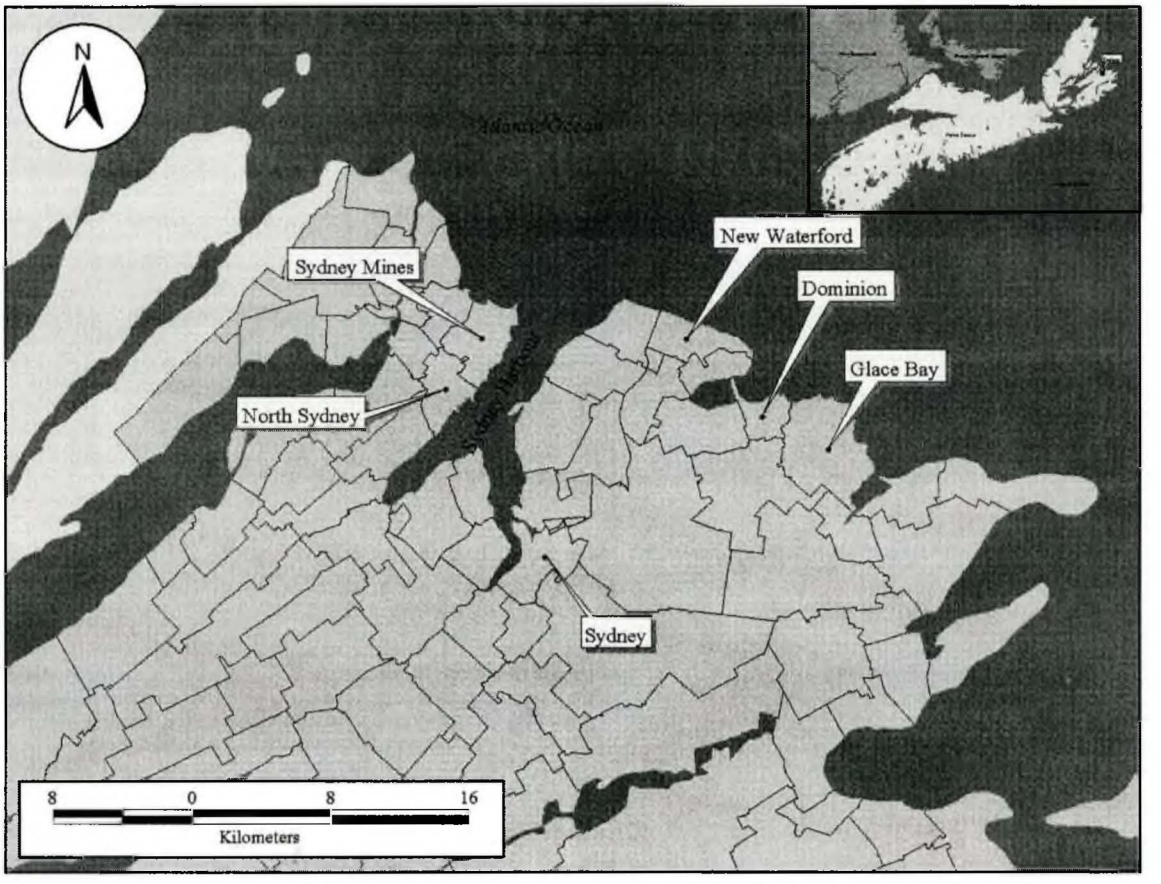


Figure 3.1. Location of Sydney, and other communities in Industrial Cape Breton.

a large plot of land on Sydney Harbour (Harvey, 1971, as cited in Burra, 2002). Full production began in 1901, with peak outputs during both World Wars (CBCL and CRA 1999, as cited in Burra, 2002). In addition to the steel plant, several allied industries were built on adjacent lands. At various times throughout the past century, the site was home to DOMTAR (a coal tar refining plant), a benzol plant, an ammonia sulphate by-products plant, a sulphuric acid plant, a brick plant, two coal washing facilities, coke and coal storage areas, and numerous tanks for tar, benzol, and other products of coke production and of course the coking facilities, just east of the steel plant (Figure 3.2). The two

bodies of water known as the Tar Ponds were constructed in the 1940s to handle the overflow of coal tar from the coking operation (Dodds and Seviour, 2001).

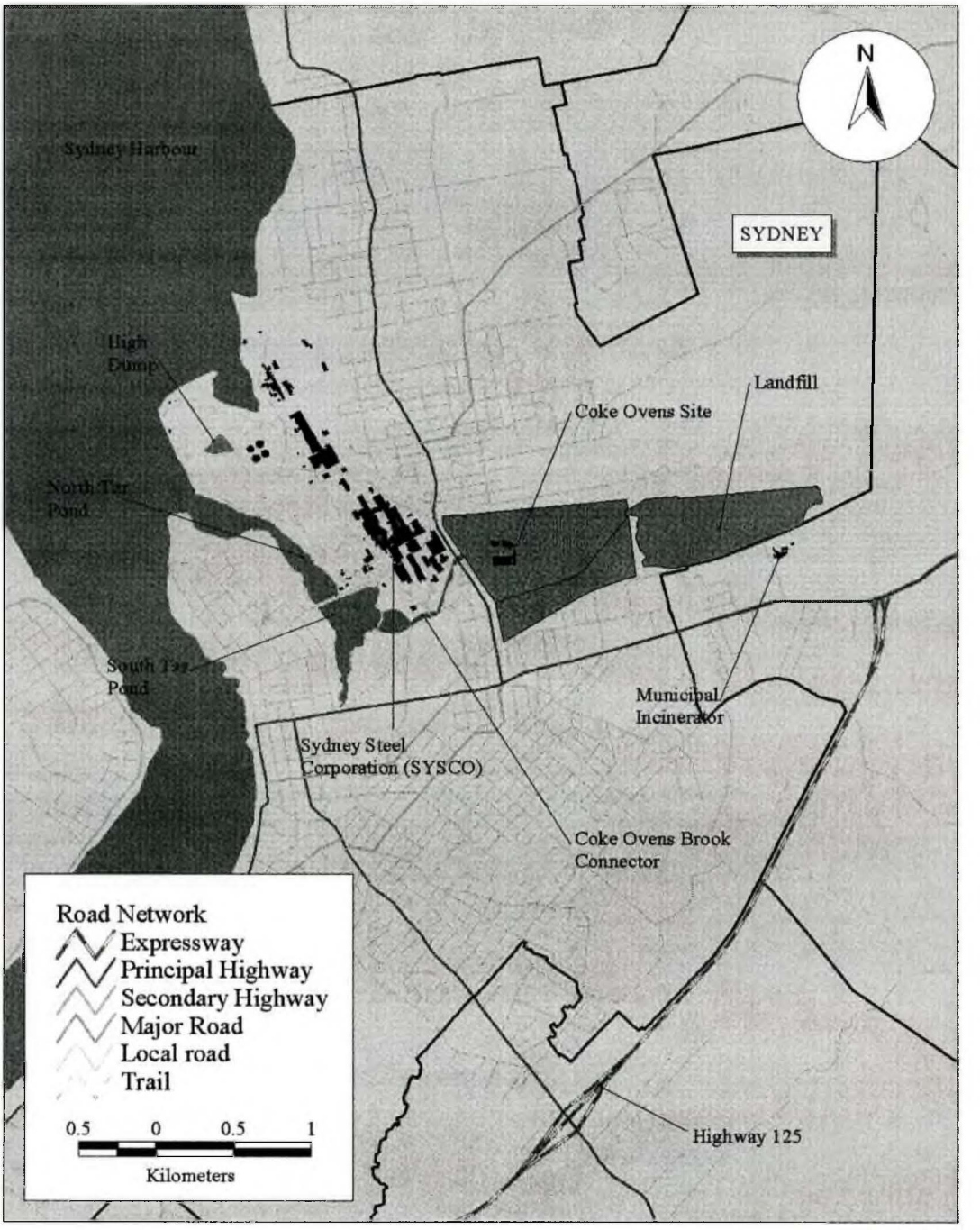


Figure 3.2. Location of point sources of pollution in Sydney.

In 1967, the steel plant was set for closure, but the provincial government intervened, and purchased the plant from the Hawker-Siddeley Corporation and renamed it Sydney Steel Corporation (SYSCO). The steel plant went through several upgrades to ensure its viability, but by 1988, with the plant not as profitable as it once was, the Coke Ovens ceased operation. By 2001, SYSCO ceased operation and by September 2002, one of the most enduring components of the Sydney landscape, the Coke Oven stacks, were dismantled piece-by-piece (JAG, 2003).

Left in the wake of the steel industry is a legacy of contamination that resulted in one of the largest industrial waste sites in North America. During operation, there were elevated levels of pollution in the air, water and soil. The air quality has improved since the coke ovens closure as indicated by monitoring carried out as part of the current remediation process (JAG, 2003). However, as various assessments demonstrate, contamination in the water and the soil remain (for example see JDAC, 2002). Current estimates have the Tar Ponds containing about 700 000 tons of contaminated sediment. Included in the sediment are various by-products of the coking process such as: coal tar, several types of PAHs, PCBs and an assortment of heavy metals. Contributing to the contamination of the Tar Ponds is a landfill site located on the eastern-most part of the property (Figure 3.2). Initially a marsh, the steel company used the area as a dump for its waste. The community soon followed suit, and in 1957, the municipality began using it as a landfill for residential solid waste. Since that time, an estimated one million tons of waste have been deposited at the site (JAG, 2003). Leachate from the landfill flows to the Tar Ponds through the Coke Ovens Brook Connector (Figure 3.2). In addition to

leachate from the landfill, millions of litres of raw sewage pour into the brook from the nearby neighbourhoods on a daily basis (JAG, 2003). The combination of chemicals and raw sewage results in a noxious odour but the contaminants are only harmful to people if they touch or ingest it (JAG, 2003). There is currently a security fence encircling the entire Tar Ponds in order to provide some security from the pollution until the remediation is completed.

3.3 Sources of Data

This section describes the types of data collected for this thesis. The following Sub-section describes the spatial data collected from various sources. This is followed by the reproductive health data collected from the Nova Scotia Reproductive Care Program (RCP). The final part of the chapter describes the census data employed in this research.

3.3.1 Spatial Data

Various types of spatial data were collected for this research. Boundary files and detailed road networks were collected from the topological map collection created by Desktop Mapping Technologies (DMTI). The Regional government in Cape Breton provided the study with a civic address point file that contains geocoded locations for every household in the municipality. The houses were geocoded by placing a dot over each home observed in an aerial photograph, thus providing a much more accurate location of residences than could be attained in a postal code conversion file (PCCF). In addition to the civic address point file, GIS data were provided for the building footprints of SYSCO steel, the municipal incinerator, the Tar Ponds, and the landfill, meaning that all of the major point sources of pollution in that area can be represented spatially on a

map. In addition, they provided a community boundary map, created by dividing CBRM into 107 communities. Finally, the HSWG of JAG (see Burra, 2002 for details) provided shapefiles of pollution plumes for the Coke Ovens and SYSCO constructed through a dispersion model.

3.3.2 Reproductive Health Data

Reproductive health statistics are routinely collected on a prospective basis by the RCP of Nova Scotia and are stored in the Atlee Perinatal Database. This database, which contains population-based data from 1988 onwards, is used for ongoing clinical audit, peer review, surveillance, and epidemiologic and clinical research. The RCP reports on a number of variables related to reproductive care and reproductive outcomes including the prevalence of congenital anomalies and stillbirths within the province, as well as incidence of low birth weights, and preterm births. Data corresponding to CBRM exist from as early as 1981, however, it is not population based as different hospitals were incorporated into the database at different times. For example, Cape Breton Regional Hospital in Sydney was incorporated into Atlee in 1985 whereas the hospitals in New Waterford and Glace Bay were incorporated in 1988. Since all observations were retrieved from the Atlee using a municipality code, it is possible to have records corresponding to residents of CBRM prior to the incorporation dates of the region's hospitals. For some of these observations, women who resided in CBRM were admitted to a hospital outside of the region that was already incorporated into the Atlee database. In some cases, these were women with high-risk pregnancies that were sent as far away as Halifax for treatment/delivery (Fahey, 2003 personal communication).

For this thesis, 294 variables for 22108 observations, spanning the years 1981 – 2002, were extracted from the Atlee Perinatal Database. Variables collected include maternal and infant demographics, and information about procedures, interventions, diagnoses and outcomes for women and newborns. A list of all the variables collected from the Atlee can be found in Appendix II. From the original list of variables obtained, several new variables were derived through either adding several existing variables together, or recoding them into dummy variables.

There are four outcomes of interest in this thesis: stillbirths, congenital anomalies, preterm births and LBW babies. Since the Atlee only stores records for post-20 week pregnancies, miscarriages are not included in this analysis since miscarriages occur prior to 20 weeks of gestation. Weinberg and Wilcox (1998) point out an important fact in relation to reproductive outcomes. That is many outcomes are not mutually exclusive in studies of this type. It is possible to have one observation coded as a congenital anomaly, a preterm birth and LBW. For the methods that will be applied throughout this thesis, it is important to have mutually exclusive outcomes. According to Wainman (2003, personal communication) there is a hierarchy that the outcomes follow. In the example above, it is likely that the congenital anomaly caused the preterm birth, and the fact that the infant was born prematurely, led to the LBW. In a case such as this, the observation is coded as a congenital anomaly. The method used for coding the outcomes is as follows. Stillbirths were derived from the 'Outcome' field provided in the Atlee (See Appendix III for a full list of outcomes). All records coded as 'FTD' (fetal death) were coded with a one to represent stillbirths, all others with a zero for live births.

With respect to congenital anomalies, all types of anomalies are recorded in the Atlee. Some of these, however, are not considered significant enough to include in what is classed as a major or minor anomaly. Typically, the malformations, which are considered major, are ones where there is an adverse effect on an individual's health, functioning or social acceptability (Forand *et al.*, 2002). The variable used for this research was obtained by summing the major and minor anomaly fields provided, into one field representing all types of major and minor anomalies. These were also coded as one for an anomaly and zero for a non-anomaly. Infants with more than one type of anomaly were coded only once.

Preterm births are defined as a delivery prior to 37 completed weeks of gestation (Kramer, 2003). The database contains a field for gestational age that was originally coded as a continuous variable, but was rounded down to the number of full weeks. This field was recoded by taking all observations with less than 37 weeks of gestation and equating them to one for all cases, and zero for all deliveries at term (post-dates were not considered for this research and were treated as term).

LBW is defined as a birth weight that is less than 2500 grams (Kramer, 2003). In this database, birth weights were originally coded as a continuous variable, however they were rounded down to the nearest 500 grams for confidentiality purposes, thus becoming a categorical variable. This means that any observation coded with a birth weight of 2000 grams can be any birth weight in between 2000 grams and 2499 grams. To indicate whether the observation corresponded to a LBW, any record coded as 2000 grams or less was recoded into a binary variable for analysis.

All outcomes were stored in separate fields in the database. Where there were observations with more than one outcome, the most dominant outcome was retained and coded as a 1 with the others coded as a 0. The order of importance was as follows: stillbirths, congenital anomalies, preterm births and LBW. This means that any observation coded as a LBW, is one where the pregnancy did not result in a stillbirth, a congenital anomaly or a preterm birth. A normal birth would have a 0 in each of the fields representing the individual outcomes.

3.3.3 Census Data

Some variables were required from the 1996 Census (Statistics Canada, 1996b) since there are not any indicators of SES in the Atlee Perinatal Database. SES has been found to be an important determinant of reproductive health as indicated in the previous chapter. Further, studies have shown lower SES to be consistently associated with adverse pregnancy outcomes (for example see Gazmararian *et al.*, 1996, Vrijheid *et al.*, 2000 and English *et al.*, 2003). To have some indication of SES, five variables were obtained from the Census and a deprivation score was created. Smith *et al.* (2001) state that deprivation scores may be used when a person's SES has not been or cannot be collected. The index calculated for this research was constructed in a fashion similar to the Townsend Index, where material deprivation is measured as the summation of the standard scores of the five variables collected. It is a relative measure that is used to rank areas, meaning an area will be either more or less deprived than other areas in CBRM.

The index, initially created for use in the United Kingdom, includes the following four variables: unemployment (lack of material resources and insecurity), overcrowding

(material living conditions), lack of owner occupied accommodation (a proxy indicator of wealth) and lack of car ownership (a proxy indicator of income) (Corporate Information Services of Devon County, 2003). In Canada, proxies for wealth and income are not necessary since these variables are routinely collected in the Census. For this research, the following five variables were used: unemployment, house in state of disrepair (material living conditions), dwelling value (indicator of stored wealth), median household income and finally, a variable for educational attainment (women without a high school diploma). These variables were collected at the enumeration area level and the index was computed. For enumeration areas with suppressed data, a spatial moving average using the data from adjacent enumeration areas was applied. To interpret this index, any positive value indicates that an area is deprived, with larger numbers indicating that the area is more deprived relative to rest of the enumeration areas within CBRM. The values were then sorted into deprivation quintiles as in Vrijheid *et al.*, (2000) for analysis.

3.4 Geocoding of the Atlee Perinatal Database

All observations in the Atlee Perinatal Database were selected by municipality codes corresponding to the communities within CBRM. Each observation had address information attached to it with varying degrees of completeness. Depending on the level of completeness, different methods of geocoding were applied to come up with x and y coordinates. For example, if a woman had provided a full address, then the postal code was used with Statistics Canada PCCF. If, however, no address information was given, but a place name was (as was the case for many of the observations prior to 1988), then

the records were geocoded to the centroid of the place. These coordinates were obtained from a Nova Scotia Gazetteer and coded in-house at IWK Grace Hospital in Halifax, Nova Scotia. Each geocode was also assigned a code indicating how the coordinates were derived for each record. Table 3.1 presents the list of codes assigned to each geocode within the Atlee database.

Table 3.1. Types of geocoding used in the Atlee Perinatal Database

Value	Description
0	Somewhere in Nova Scotia (geocoded to the centroid of the province)
1	Urban postal code; unique latitude and longitude
2	Urban postal code; multiple latitude and longitude in small area
3	Urban postal code; unique latitude and longitude probably not equal to municipal code
4	Urban postal code; multiple latitude and longitude in a larger area
5	Urban retired postal pode
6	Urban retired postal code; multiple latitude and longitude in a small area
7	Gazetteer
8	Gazetteer using place name and county
9	Municipal code
10	Municipal code (rural)
11	Rural or dubious postal code

The PCCF from Statistics Canada has several limitations (for full details, see Statistics Canada, 1996c). There are 2 limitations that are particularly relevant for this database. The first is that several of the postal codes correspond to rural areas, which is quite problematic since the PCCF will put several events in the same location. The problem is that any rural postal code covers an area that is quite large and it usually corresponds to a Post Office Box, which services far-reaching areas (Burra *et al.*, 2002). The second problem lies in the placement of the geocode within urban areas. In these

areas, the geocode is placed on the most representative block face within an enumeration area (Statistics Canada, 1996c, Burra *et al.*, 2002). Although enumeration areas can be quite small in urban areas, there may still be several different postal codes geocoded to the same location. If, for example, there are five postal walks within the enumeration area, all five postal codes for those walks will have identical x and y coordinates. These coordinates will correspond to the largest, or most representative postal walk within the enumeration area. For example, in Whitney Pier, there are 37 valid postal codes. When using the Statistics Canada's PCCF, only 7 dots are plotted on a map (one per enumeration in Whitney Pier), meaning that multiple postal codes exist at each location. To correct this problem, the following algorithm was applied to all records to improve the geocoding:

1. In order to achieve the highest level of accuracy possible, address geocoding was attempted by joining the civic address point file provided by CBRM to the records with full addresses. For any records that fail to match (due to spelling errors, wrong street types, directions etc.) the original information was retained (i.e., the postal code and the original x and y coordinates from StatsCan's PCCF).
2. All records not geocoded in step one but with a valid postal code were geocoded using the PCCF from DMTI. This value-added PCCF provides much more detail since DMTI geocodes each postal code to the centroid of each block face as opposed to the most representative block face within the enumeration area. This means that each postal walk is now represented spatially. In Whitney Pier, for example, there are 37 unique pairs of x and y coordinates, one for each of the

postal walks, as opposed to 7 unique pairs of coordinates in Statistics Canada's PCCF. For any records that are not successfully geocoded from DMTI, the original coordinates were retained.

3. All data not geocoded in the first two steps that only have latitude and longitude information or a fictitious postal code acting as a placeholder in the database, were separated from the rest of the records and the coordinate information was retained. These records correspond to coordinates assigned from the Gazetteer, and no improvement is possible.
4. All records corresponding to the Eskasoni or Membertou Indian Reserves were extracted since there is no detailed address information to match each observation (street names within the reserves are not available to the general public). The original coordinates were retained.

Once all of the geocoding was completed, all records were merged into one large file, and all records falling outside of CBRM were removed (n=721) leaving 21387 valid records for analysis.

3.5 Analytic Methods

This section details the analytic methods employed in this thesis. The first section outlines the ESDA applied including methods of visualization, exploration and modeling of marked point patterns. This is followed by the categorical analysis employed in this research.

3.5.1 Exploratory Spatial Data Analysis

Spatial analytical techniques for point patterns were applied to the data. While useful tools when studying reproductive outcomes (Dolk, 1999, Gatrell, 2002), few reproductive studies using these techniques exist (Rushton and Lolonis, 1995, Dolk *et al.,* 1998b, Reader, 2001, Gatrell, 2002) due to the lack of individual level data. Moreover, point pattern techniques are most useful with a full enumeration of the population rather than a sample, so the full value of the analysis cannot be attained with data from a cross-sectional survey (e.g., Burra, 2002). This technique is possible, however, using the Atlee database given it is population based from 1988 onwards. In the present study, the reproductive outcomes form a multivariate point pattern, with each type of outcome being represented spatially in what is known as a marked point pattern. The typical question asked when dealing with any point pattern is whether it exhibits any form of clustering or regularity. The usual method to determine this is to compare the observed pattern(s) to a theoretical pattern of complete spatial randomness (CSR). Under CSR, all events are distributed independently over the study region and the observed pattern is compared to the theoretical pattern to determine if there are any departures from randomness in either direction (i.e., clustering or regularity). As Gatrell *et al.* (1996) point out, when studying phenomena related to humans, CSR loses its usefulness. The reason for this is that it is known *a priori* that there is built in heterogeneity in the point pattern due to the underlying population density. It is known that all cases of adverse reproduction must occur in areas where there is an existing population, so there will always be a departure away from CSR and towards clustering when studying phenomena

of this type. It is critical then to correct for the underlying population in any form of spatial analysis.

The section that follows presents the methods adopted to create and visualize the point pattern data set. Visualization is always the first step in any form of spatial analysis. Following this, two advanced exploratory/modeling techniques are presented, where the underlying population has been accounted for. The aim of this phase of the analysis is to address the spatial pattern of the outcome data.

3.5.1.1 Visualizing Negative Reproductive Events

Point patterns were created for each outcome under study using the entire geocoded Atlee database (N=21387) as a starting point. Since the geocodes prior to 1988 were poorer in quality, they were removed for the spatial analysis (n=2254). Additionally all non-singleton fetuses were removed (n=392), since the prevalence of negative events is greater among non-singleton births (Forand *et al.*, 2002). Finally all records corresponding to residents of the Eskasoni community (n=1053) were removed since they are First Nations (Mi'kmaq). These records had to be removed since they did not have any detailed geographic information and all observations were coded to the same location. Additionally, information from the census is not available for First Nations. From the remaining points (n=17648), dot maps were created for each of the negative outcomes (see Chapter 4) covering the entire boundary of CBRM. The next section describes the exploratory analysis that follows visualization.

3.5.1.2 Bivariate K Functions

The K-function is the best method available to assess the second-order effects in the dataset and it provides a description of the level of autocorrelation present in the point pattern. In other words, the K-function seeks to answer the question; does the presence of a negative reproductive event make it more or less likely that another negative reproductive event will be found nearby? This statistical method aids in addressing the first research objective; to assess the spatial pattern of various types of adverse reproductive events plausibly linked to the environment exposure of interest.

The K-function is defined as the expected number of events within a specified distance of an arbitrary event. It is calculated via an estimator visiting each event in the point pattern and counting all events at several spatial lags around the event in question, as defined by the user (see Bailey and Gatrell, 1995 for a full description). Under the null hypothesis, there is an absence of second order effects meaning that there is a lack of spatial autocorrelation (i.e., an absence of clustering). As mentioned previously, given the heterogeneity present in the spatial distribution of the underlying population, the expectation is that the K-function will indeed indicate clustering, rendering the null hypothesis meaningless. The alternate hypothesis of clustering will almost always be accepted, since the majority of negative reproductive events will be located in the major population centers, where there is an abundance of total reproductive events. The interest, then, needs to shift to clustering above and beyond what is expected based on the existing population distribution.

Diggle and Chetwynd (1991) have proposed a test based on the K-function that takes the underlying population into account and addresses the overall global pattern in the data set. They propose to calculate two separate K-functions, one for the cases, and one for a representative set of events acting as controls, and to calculate the difference between them. When the difference is plotted over all spatial lags, it can be determined where the pattern of cases is more clustered than the pattern of controls. So, for the cases, the K-function is defined as:

$$K_{c,c}(d) = \lambda_c^{-1}\text{E[number of additional cases} \leq \text{distance } d \text{ of a randomly chosen case]} \quad (3.1)$$

where λ_c is the density of cases and E[.] is the expected value of the expression in the square brackets (Haining, 2003). The K-function for the controls $K_{k,k}(d)$, is defined in a similar manner. The bivariate K-function (calculated for the marked point pattern) can then be stated as:

$$K_{c,k}(d) = \lambda_k^{-1}\text{E[number of controls} \leq \text{distance } d \text{ of a randomly chosen case]}. \quad (3.2)$$

The underlying hypothesis to be tested here is not one of independence, but rather one of random labeling. Under the random labeling hypothesis, each pattern (cases and controls) taken separately is considered to be a random "thinning" of the marked point pattern, which, in this case, is the population. Bailey and Gatrell (1995) point out that K-functions are invariant under random "thinning", so under the random labeling hypothesis it is theoretically expected that:

$$K_{c,c}(d) = K_{k,k}(d) = K_{c,k}(d). \quad (3.3)$$

So the difference between the two K-functions is a sensible test to detect departures from random labeling. The test statistic, given by Haining (2003) is defined as:

$$\hat{D}_{c,k}(d) = \hat{K}_{c,c}(d) - \hat{K}_{k,k}(d) \tag{3.4}$$

If the distribution of cases is the same as the distribution of controls, the result of equation 3.4 is zero and is represented graphically as a horizontal line when $D_{c,k}$ is plotted against distance. Values above zero will be represented graphically as peaks, and correspond to spatial clustering beyond the background population distribution. To determine if the peaks are significantly different from horizontal, Diggle and Morris (1996) suggest applying Monte Carlo simulation to create upper and lower simulation envelopes. The process used in generating the simulation envelopes is to take the marked point pattern and randomly assign the cases, where the total number in each simulation is equal to the total number of cases in the data set. In other words, the point pattern of the cases from the dataset is combined with the point pattern of the controls to create the marked point pattern. For each simulation, the locations of all combined events are held constant, and the cases are randomly assigned such that the new numbers of "cases" are equal in size to the number of cases in the data set. For each new simulation, the difference of K-functions is calculated at each distance, and the maximum and minimum values are retained to create simulation envelopes. If the empirical value of the test statistic exceeds the maximum simulation envelope for any distance *d*, this indicates a clustering of cases at that distance. If the empirical value of the test statistic is less than the lower simulation envelope, then the presence of a case inhibits the presence of other cases at distance *d*.

This statistic has been applied successfully to cases of childhood leukemia and lymphoma in North Humberside (Diggle and Chetwynd, 1991), and leukemia in Lancashire (Gatrell *et al.*, 1996). Two examples dealing with a reproductive outcome were found. The first was in Gatrell (2002), where this test was applied to congenital heart malformations in Lancashire and Cumbria, 1985-1994, however, this work has not been peer-reviewed. The second was in Dolk *et al.* (1998b) where the test was applied to cases of congenital anophthalmia and microphthalmia.

For this research, individual point patterns were created for each of the outcomes mentioned previously. Controls were selected from all normal live births as coded in the Atlee database. They were selected by choosing the two normal pregnancies closest in time to the negative event being examined. This usually corresponded to the normal pregnancy immediately before and immediately after the negative outcome. Controls are usually chosen in this fashion to account for seasonality, a possible difference in exposure over time and the fact that new knowledge and advances in medicine may arise over time. Dolk *et al.* (1998b) used three controls for each case in their study, but they did not describe the selection procedure. Gatrell (2002) however, points out that it is quite sensible to select the two normal births closest in time to the case to act as controls. If, for example, a case occurred at the beginning of the time period, it would make little sense to select two controls that occurred 15 years later. The difference of K-functions was calculated based on the event and its respective control pattern over a 25 kilometer radius. Significance was determined using Monte Carlo simulation with 99 simulations being computed. This statistic was computed for the entire 15-year period (1988-2002) as

well as for three five-year periods (1988-1992, 1993-1997, 1998-2002). These were used since it was possible that a signal may have been detected in one of the time periods, which may not have been detected using the entire 15 years. All calculations were done using S-Plus 2000 from Insightful Corporation and the Splancs library (Rowlingson and Diggle, 1994)

3.5.1.3 Ratio of Kernel Estimates

While the difference of K functions is a useful test to assess the overall spatial pattern of negative reproductive outcomes in CBRM, it does not identify where individual clusters occur. To determine where clusters might be located, the first-order effects need to be examined. These effects correspond to a variation in the mean of the process over the study area. The most effective way to study the first-order effect in this dataset is through kernel estimation.

For any given point in the study area, kernel estimation provides the number of events per unit area at that point. Operationally, the intensity or density of the point pattern is assessed through an estimator visiting several user defined locations on a grid, and assessing the number of events that fall within the radius of the estimator. The shape of the estimator can vary, but usually it is based on the quartic kernel. The quartic kernel (ignoring edge-correction) is given by:

$$\hat{\lambda}_{\tau}(s) = \sum_{h_i \leq \tau} \frac{3}{\pi \tau^2} \left(1 - \frac{h_i^2}{\tau^2}\right)^2 \tag{3.5}$$

where τ is a smoothing parameter known as the bandwidth and h_i is the distance between the point s and the observed event location s_i. The bandwidth is user defined and the

shape and radius remain fixed as the estimator moves through the grid. There are extensions to the method that allow the bandwidth to vary, such that the bandwidth is narrower in high-density areas and wider in less dense areas. In either case, the numbers of events within the radius are weighted such that the events closer to the origin of the estimator have more weight than those further away (see Bailey and Gatrell, 1995 for further details). Forming estimates on a grid is preferable to the more standard ways of mapping incidence by discrete boundaries since these areas are likely to change over time, and rates do not drastically change due to a border (Rushton, 2003). As with the K function, the standard formulation of the kernel is not adequate when dealing with human phenomena since it is known *a priori* that clustering will occur in the major population centers.

Clusters of adverse outcomes were corrected for the underlying population as in Bailey and Gatrell (1995). Since the phenomena under study are dependent on the location of humans in space, one would expect to find more negative cases occurring in areas where there is a larger population and therefore more pregnancies. To account for this, the underlying population is controlled for through a ratio of kernel estimates using equation 3.6:

$$\hat{\rho}_{\tau}(s) = \frac{\sum_{i=1}^{n} k\left(\frac{s - s_i}{\tau}\right)}{\sum_{j=1}^{m} k\left(\frac{s - s'_j}{\tau}\right)} \tag{3.6}$$

where k() is the quartic kernel, and the edge-correction has been ignored for simplicity (see Bailey and Gatrell, 1995 for formulation). In a study dealing with the spatial

distribution of Ontario residents attending McMaster University, Buliung and DeLuca (2000) have demonstrated marked differences in the patterns that emerge when the correction is applied. In CBRM, it is well known that most of the population resides in the former municipalities mentioned earlier in this chapter; hence, the expectation is that the majority of events occur there as well. For this research, the study area was limited to a 20-kilometer radius around the Tar Ponds. This distance captures the areas that comprise ICB, and eliminate much of the rural areas to the south of Sydney, where the points are sparsely distributed. The numerator of the ratio is generated by the events (stillbirth (n=100), congenital anomalies (n=748), preterm births (n=1453), and low birth weights (n=168)). The denominator is a kernel based on either all births (n=16290) or live births (n=16190) (depending on the outcome under study). This will then show the population at risk, expressed as events per 1000 births, and how it varies over space. Again, for this phase of analysis, only data corresponding to years post-1988 were used since prior years lacked sufficient spatial information.

Several constraints or filters were applied to the estimation process to produce more reliable results. The first constraint is that all grid cells in the denominator that have a value of less than one birth per square kilometer were removed. This is essential since dividing the numerator by a value less than one will greatly inflate the resulting estimate. Second, all grid cells that fall in the water bodies were eliminated. Third, in the final grid (the ratio of kernels), all density estimates less than 0.01 were removed. This equates to 10 events per 1000 births. The rationale behind this is that each grid cell in the estimation will have some value calculated by default. In many cases this value will be 0 or

something quite small, so to eliminate the background noise, these estimation points are removed. In all cases, a quartic kernel was applied with a fixed bandwidth of one kilometer. All calculations were carried out in CrimeStat II (Levine, 2002)

Building on work by Reader (2001), significant clusters were determined using random labeling and Monte Carlo simulation. In essence, a distribution is generated through simulation and we compare each grid cell in the observed case to the average of all simulations. The simulation process is as follows:

1. Using the appropriate set of data, (i.e., either all births, or all live births) there was a random assignment of cases to *n* records, where *n* is the same number as the outcome under study. This is done under the random labeling hypothesis with the same rationale as for the bivariate K-function. Using this hypothesis, the assumption is made that any of the records in the full data set are possible locations for an adverse outcome to occur.
2. Calculate the ratio of kernel estimates as above for the new point pattern using the same parameters as for the observed case.
3. Repeat the process 99 times

Upon completion of the simulation process, the observed number of events per 1000 births was compared to the simulated number of events per 1000 births to determine where there was a statistical difference. Significant differences were determined when the observed was at least 2 standard deviations away from the mean of the simulations (see Chapter 4).

3.5.2 Categorical Data Analysis

The ESDA described in the previous sections are useful for addressing the first research objective (i.e., to assess spatial patterns). In part, ESDA also helps to address the second objective in that it can be determined if there is a spatial relationship between the Tar Ponds/Coke Ovens site and clusters of negative outcomes. To model the impact of proximity to the Tar Ponds on negative outcomes, as well as to explain the incidence of the outcomes using the predictors obtained from the Atlee Perinatal Database, more advanced methods are required. In this thesis, the dependent variable was measured at a nominal scale; as such, statistical tests requiring the dependent variable to be normally distributed are not appropriate for the analysis. Analysis was first conducted on a binary variable representing each negative outcome separately. Following this, analysis was conducted on all outcomes coded as a multinomial variable. The data used for the remainder of the analyses was a subset of the Atlee database, and was different from the set used in the previous sections. For ESDA, in general, all that is required is sufficient information to correctly identify the outcomes and their location (see section 3.3.2). For this phase of the analysis, emphasis is not solely based on location. Interest now has shifted towards being able to identify what gives rise to the negative outcomes explored in the previous sections.

Similar to the subset used for ESDA, all records falling outside of CBRM were removed (n=721), as well as all non-singleton births (n=440) and all records corresponding to Eskasoni (n=1114). In addition to these records being removed, all observations with missing data for predictors such as smoking (n=823), marital status

(n=401), gestational age (n=236), and information on previous pregnancies (n=400) were removed. If information regarding birth weights or gestational age were missing, the observations were deleted unless they were an anomaly or a stillbirth, in which case they were coded as such. All observations with a missing maternal identifier were deleted (n=176) since it was not possible to examine multiple pregnancies from the same woman. This is important since reproductive outcomes are dependent on reproductive history. Finally, all records prior to 1988 were eliminated due to the poor accuracy of the spatial information. From the remaining observations (n=15800), the analysis proceeded using the full data set and then it was repeated using a data set restricted to observations where parity was zero, gravidity was one and abortions (spontaneous or induced) were zero (n=6009). By restricting the analysis to the maternal first pregnancy (primiparity), the dependence on reproductive history was removed from the analysis (Lin *et al.*, 2001;Yang *et al.*, 2002). Once all of the appropriate records were obtained, frequencies were run on the maternal identifier field to ensure they were all representing unique women. Where duplicate records appeared (signaling a coding error for parity, gravidity and/or abortions) the most recent record was removed, leaving the first pregnancy. The remainder of this chapter is divided into sections covering the descriptive analysis as well as the multivariate methods applied to this dataset.

3.5.2.1 Descriptive Analyses of Adverse Reproductive Outcomes

The first phase of the categorical analysis was descriptive in nature. For each of the outcomes in the analysis, frequencies were calculated (see Chapter 5). This was followed by bivariate analyses (see chapter 5). The primary method applied in this phase

was contingency analysis (crosstabulations) and was carried out in SPSS 12 (SPSS, 2004). In conjunction with the contingency analysis, Chi-squares and Odds Ratios were calculated where appropriate. Contingency analysis was performed first with several geographical variables derived in a GIS. Several distance bands were created using a GIS to determine if the odds of having an adverse reproductive outcome increased with proximity to the Tar Ponds/Coke Ovens site. Distance bands of 0-1, 1-2, and 2-3 Km were created. In addition, the total distance (i.e., 0-3 Km) was examined. Several studies have employed this method in the past (see Bhopal *et al.,* 1999 and Dolk *et al.,* 2000 for example). Contingency analysis was also performed using the outcomes and three zones as defined by Guernsey *et al.* (2000). In their work, they were interested in examining differences between people residing in Sydney, ICB excluding Sydney and then the remainder of CBRM. Finally, straight-line distance was also examined (as in Dolk *et al.,* 2000). To determine if there was a significant difference in distance to the site, a paired t-test (assuming inhomogeneous variances) was applied to each of the adverse outcomes and the normal births.

Following the contingency analysis with the geographic variables, the same type of analysis was repeated with the variables obtained/derived from the Atlee Perinatal Database (see Table 5.2 for a complete list and variable definitions). All significant associations were then combined for multivariate analysis.

3.5.2.2 Multinomial Logistic Regression Models

When modeling adverse reproductive outcomes, it is important to note that a pregnancy has several possible outcomes. As such, the problem is inherently

multinomial in nature, and any modeling could make use of a multinomial dependent variable reflecting the types of outcomes that may be experienced. The best way to model a problem of this nature is through a multinomial logistic (MNL) model. In general, logit models are used to model a relationship between a categorical dependent variable and one or more independent variables, where the dependent variable comes from a set of mutually exclusive categories. These methods are used to describe and/or predict discrete choices of decision-makers, or to classify a discrete outcome according to a host of regressors (TRB, 2003). In this instance, the dependent variable was the list of reproductive outcomes described earlier. MNL models are multi-equation models. A response variable with **k** categories will generate **k-1** equations. In this case, there were five categories (normal birth, stillbirth, anomaly, preterm birth and LBW), meaning that 4 equations were generated. Each of these four equations is a binary logistic regression comparing a group with the reference group (normal pregnancies). MNL regression simultaneously estimates the four logits. Further, it is also the case, that the model tests all possible combinations among the five groups although it only displays coefficients for the four comparisons.

The most frequent use of this type of modeling in a geographical setting incorporates a discrete choice framework, whereby the analyst may wish to model the choice of automobile purchase (from a set of vehicle classes) (Train, 1993), the choice of travel mode (walk, transit, rail, auto, etc.)(Ben Akiva and Lerman, 1985), or land development choice (high-density residential, commercial, suburban, etc.)(Mohammadian *et al.*, 2002). In this study, there is no decision or choice to be

made; there is an outcome that is experienced by each individual in the data set and there are several regressors that will give rise to the outcome that is experienced. With respect to studies revolving around environment and health, the discrete choice framework is not common, however, mathematically, choice models and event classification models are equivalent (TRB, 2003), so it is possible to model reproductive outcomes in this way. The response variable in this research meets the three assumptions as described by Ben Akiva and Lerman (1985). The first assumption is that the set of choices or classifications must be finite. In this case, the universal choice set C contains five elements (normal births, congenital anomalies, low birth weight, stillbirth and preterm birth), corresponding to the five broad types of events that may occur post-20 weeks gestation. The second criterion is that the set of choices or classifications must be mutually exclusive. That is the case in this research as a birth is classified into one of the aforementioned discrete categories. The final criterion is that the set of choices or classifications must be collectively exhaustive. This requirement is met as well, since any post 20-week pregnancy must result in one the categories mentioned previously.

The results from the contingency analysis were used to MNL regression models to assess the importance of several risk factors in explaining the occurrence of each of the negative reproductive outcomes. Models were computed for the full data set as well as for the subset of primiparity cases. All variables significant at the alpha level of 0.1 were entered into the model and were included in one block for this analysis. Insignificant variables ($p > 0.05$) were eliminated in backward stepwise fashion starting with the variables with the highest probability levels.

All models were assessed using the ρ^2 statistic where ρ^2 equals one minus the ratio of the maximized log likelihood for the fitted and constant only models (Agresti, 1996). Values ranging from 0.2 to 0.4 represent a very good fit of the model (Wrigley, 1985).

3.6 Summary

This chapter described the study design undertaken to meet the research objectives outlined in section 3.1. A database was obtained from Nova Scotia RCP, cleaned and geocoded using a variety of approaches to assign spatial information to the observations (Sections 3.3 and 3.4). The data were then examined using ESDA to assess whether there was a tendency for like-cases to cluster in Sydney and the surrounding area. Following this a variety of analytical methods were used to analyze these data (Section 3.5). The results of these analyses are presented in chapters four and five.

CHAPTER FOUR

RESULTS OF EXPLORATORY SPATIAL DATA ANALYSIS

4.1 Introduction

This chapter presents the results of the ESDA carried out in this thesis. Multivariate analyses are discussed in chapter five. In this chapter, ESDA was used to address the first of the research objectives:

i. To assess the spatial pattern of various types of adverse reproductive events, plausibly linked to the environmental exposure of interest;

With respect to the first objective, several point patterns were generated (one per negative outcome) as a result of the geocoding of the Atlee Perinatal Database in order to visualize the process in CBRM. This was followed by two methods used to explore and model the spatial pattern. The first method was an extension of the K-function (Bailey and Gatrell, 1995) where the interest was whether an adverse reproductive event was more clustered than a suitable control. This statistic provides an indication of whether there is global clustering present in the data. To assess the location of specific clusters, an extension of kernel estimation with a correction for the underlying population distribution was applied. These two methods together, aid in assessing the overall spatial pattern, and in part helped to determine if the observed pattern was related to proximity to the Tar Ponds/Coke Ovens site.

4.2 Geocoding and Visualization of Adverse Reproductive Events

This section describes the results of the geocoding process applied to the observations in the Atlee Database. Following this, dot maps are presented for each of the outcomes under study.

4.2.1 Geocoding

As mentioned in chapter three, all observations in the Atlee Perinatal Database were selected by municipality codes corresponding to the communities within CBRM. Each observation had address information attached to it with varying degrees of completeness and based on this information, were geocoded in-house at IWK Grace Hospital in Halifax. Due to the limitations described in chapter three, an algorithm was applied to improve the accuracy of the geocodes. These steps produced geocodes of four types. The highest levels of accuracy were attained by address matching the observations with a civic address point file provided by CBRM. The remaining records that did not match up with an address were geocoded using a PCCF provided by DMTI Spatial. The original coordinates from Statistics Canada's PCCF were retained for any of the observations that could not be matched in the previous two steps. Table 4.1 shows the numbers of records successfully geocoded at each step. Observations corresponding to Eskasoni (n=1093) were separated since there is no detailed road or postal code information for them. As a result, the original coordinates were retained.

While it may be argued that using different methods to geocode could introduce an additional form of bias or error, there was no indication that the observations of primary interest (i.e., the adverse outcomes) were geocoded in a manner systematically

different from the normal pregnancy outcomes. The only systematic difference in the geocoding occurred with respect to the year of the observation. However, both normal and adverse pregnancies had equally poor positional information prior to 1988.

Table 4.1 Results of the Geocoding Process

Type	Number of Successful Geocodes	Percentage of Data Set (%)
Address Matching	3008	17
PCCF from DMTI	14410	80
Original Coordinates	539	3

Upon completion of the geocoding process, the reproductive events were mapped, and all records outside of CBRM's boundaries were removed (n=721). Additionally, all records prior to 1988 (n=2254) were removed for reasons outlined previously. An additional 392 cases had to be excluded from consideration since they were non-singleton births. Finally, 1093 observations pertaining to Eskasoni were eliminated. From the remaining 17648 observations, dot maps were created to display the locations of the various adverse outcomes.

4.2.2 Dot Maps

Several dot maps were created once the geocoding was completed. Maps of this nature proved useful, in that any points that fell outside of CBRM's boundaries were identified (n=721) and subsequently removed. Additionally, maps of this type are quite useful to assess any broad patterns across the study area. Figure 4.1 presents the distribution of all births in CBRM. From this figure it is noticeable that the majority of the events occur in the northern portion of the county where the majority of the

population lies. Specifically, births appear to be most clustered in the areas of ICB. In addition to these places, there appears to be a large number of events occurring in the communities immediately surrounding Sydney (i.e., Coxheath, Sydney River and Westmount). There is also a smaller grouping of events in Louisbourg. Events to the south of the Sydney area and the rest of ICB are spread throughout the remaining regions, reflecting the underlying population density of the region.

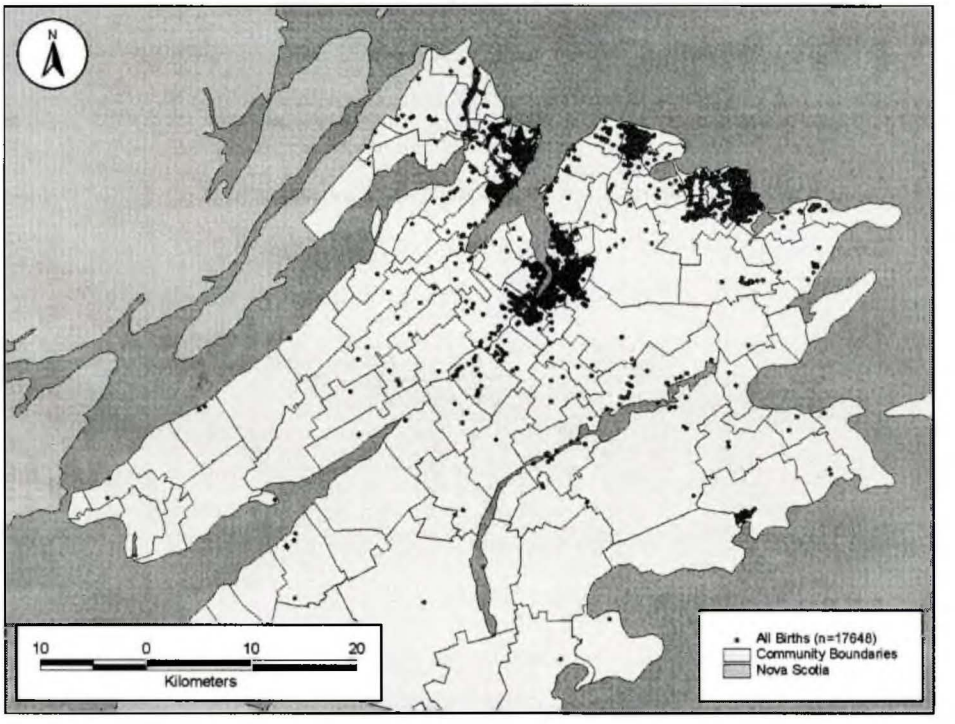


Figure 4.1. Distribution of all births in CBRM, 1988-2002 (n=17648).

The distributions of the adverse outcomes (preterm births, congenital anomalies, LBW and stillbirths) represent subsets of all births, and as such, mimic the pattern of all births, where the majority of the observations are in ICB. The largest adverse event is Preterm births (n=1604). The second most prevalent event type is congenital anomalies

(n=835) and is followed by LBW (n=185) and finally, stillbirths (n=100). All of these patterns are presented in Figures 4.2 through 4.5 respectively.

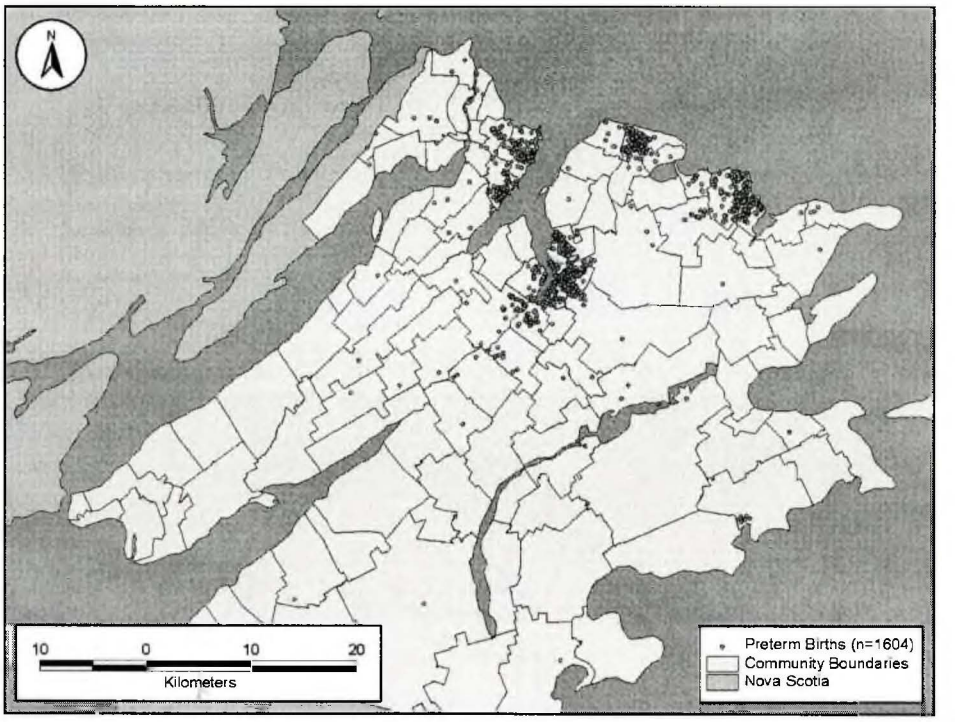


Figure 4.2. Distribution of preterm births in CBRM, 1988-2002.

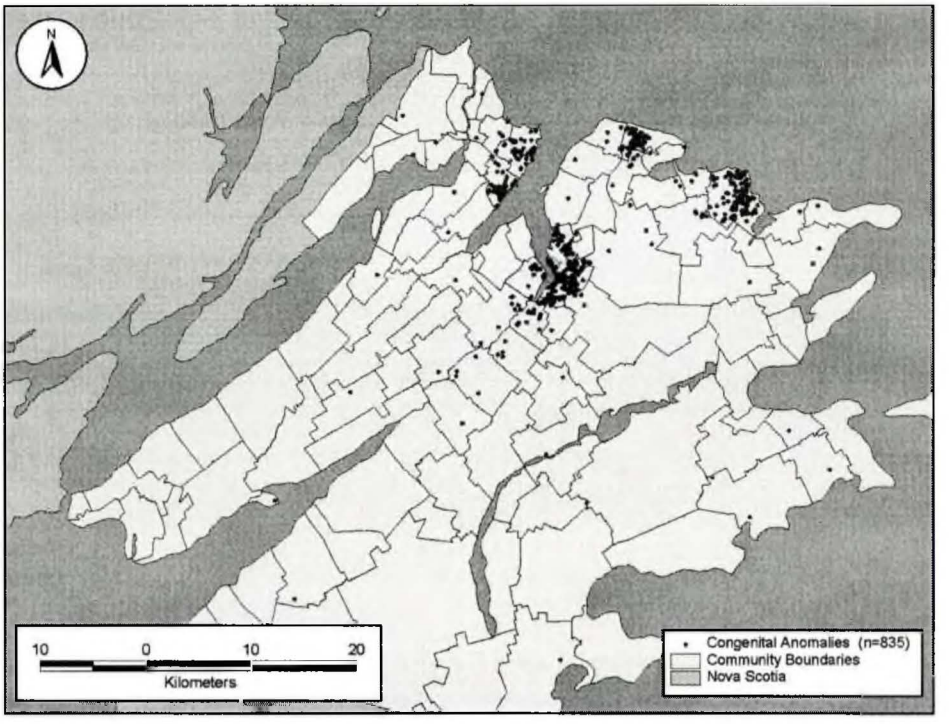


Figure 4.3. Distribution of congenital anomalies in CBRM, 1988-2002.

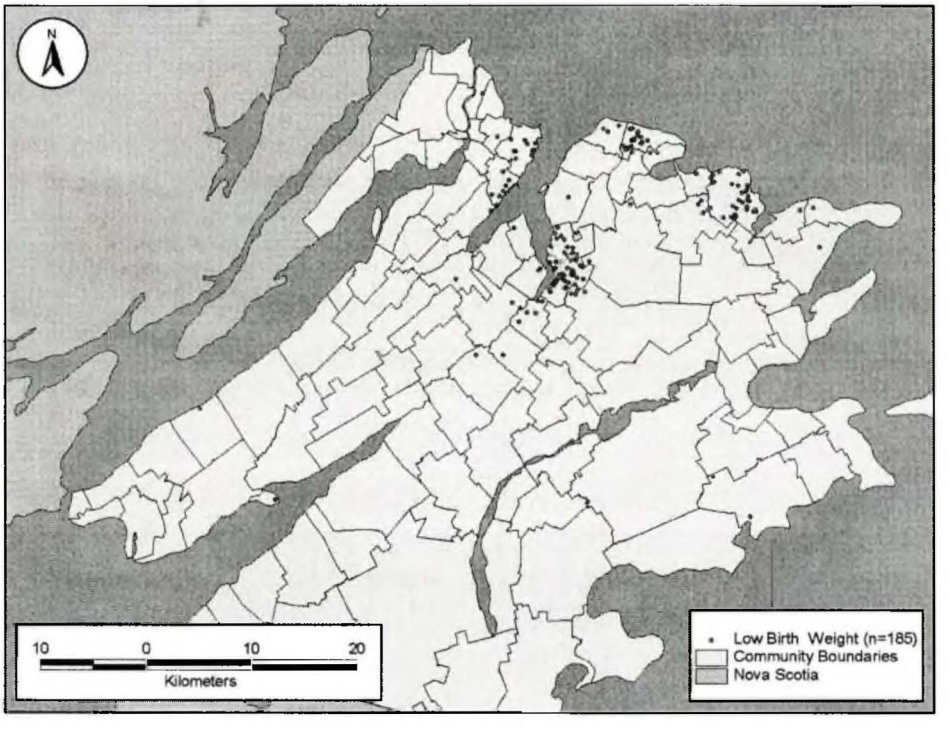


Figure 4.4. Distribution of LBW in CBRM, 1988-2002.

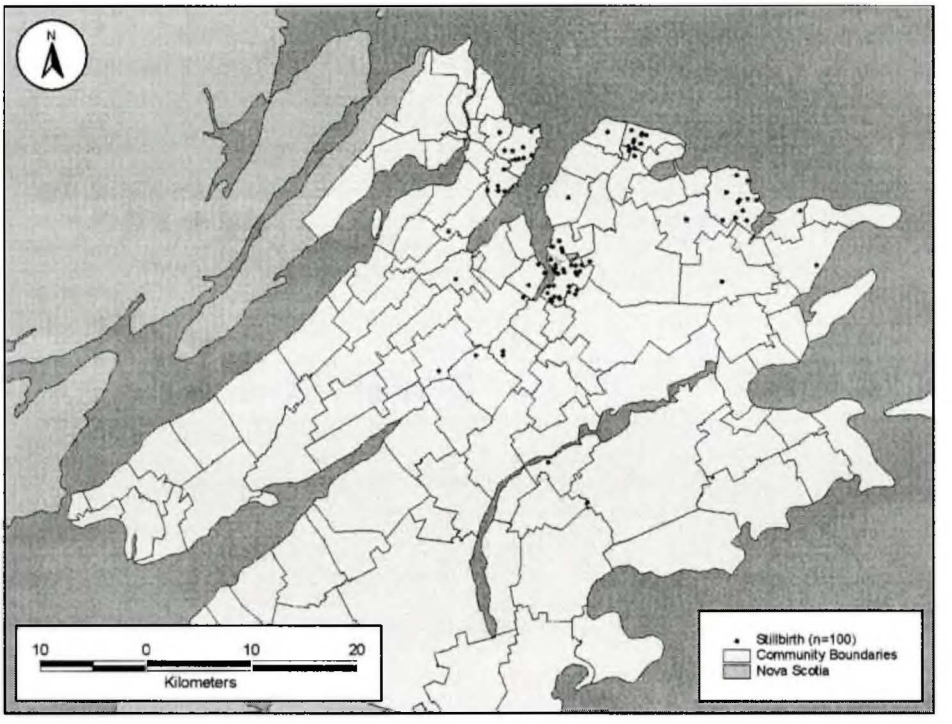


Figure 4.5. Distribution of stillbirths in CBRM, 1988-2002.

4.3 Exploration and Modeling of Spatial Point Patterns

The visualization stage of the analysis was useful in that it displayed what appeared to be clusters of events in the northern part of CBRM. Further, through the use of dot maps, it was demonstrated that events to the south of Sydney appear to be dispersed and that many communities lacked events altogether. As a result of the visualization it was determined that part of the spatial analysis (ratio of kernel estimates) should be conducted using a 20-KM radius around the tar ponds (since areas to the south were sparsely populated). This area covered all of the major communities in ICB. Prior to this step, however, there was a need to determine if there was any clustering on a global scale in the data set. As mentioned in chapter three, given the heterogeneity present in the spatial distribution of the underlying population, the expectation is that any

spatial measure will provide an indication of clustering in a data set dealing with human phenomena. As a result, the interest is whether there is clustering above and beyond what is expected based on the underlying population distribution. The test applied to examine clustering on a global scale was the bivariate K-function.

4.3.1 Bivariate K-Functions

In order to apply the method proposed by Diggle and Chetwynd (1991), two point patterns were required, a pattern of cases and a pattern of controls. The point patterns used in the visualization phase of the analysis represent the cases. Controls were selected from all normal live births as coded in the Atlee database. They were selected by choosing the two normal pregnancies closest in time to the negative event being examined. This usually corresponded to the normal pregnancy immediately before and immediately after the negative outcome. The difference of K-functions was calculated at 25 meter intervals (up to 25 km) based on the event and its respective control pattern. Significance was determined using Monte Carlo simulation with 99 simulations being computed. This statistic was computed for the entire 15-year period (1988-2002) as well as for three five-year intervals (1988-1992, 1993-1997, 1998-2002). Five-year periods were used as well since it is possible that a signal may be detected in one of the time periods, which may not be detected using the entire 15 years. The output of this statistic is graphical in nature, and its interpretation is simple; wherever the observed difference between the K-function for the cases and the K-function for the controls is greater than what is expected through simulation (i.e., the observed statistic falls outside of the simulation envelope), then there is a significant difference between the two patterns.

Figure 4.6 shows the results of the difference in K-functions for preterm births for the entire period (1988-2002).

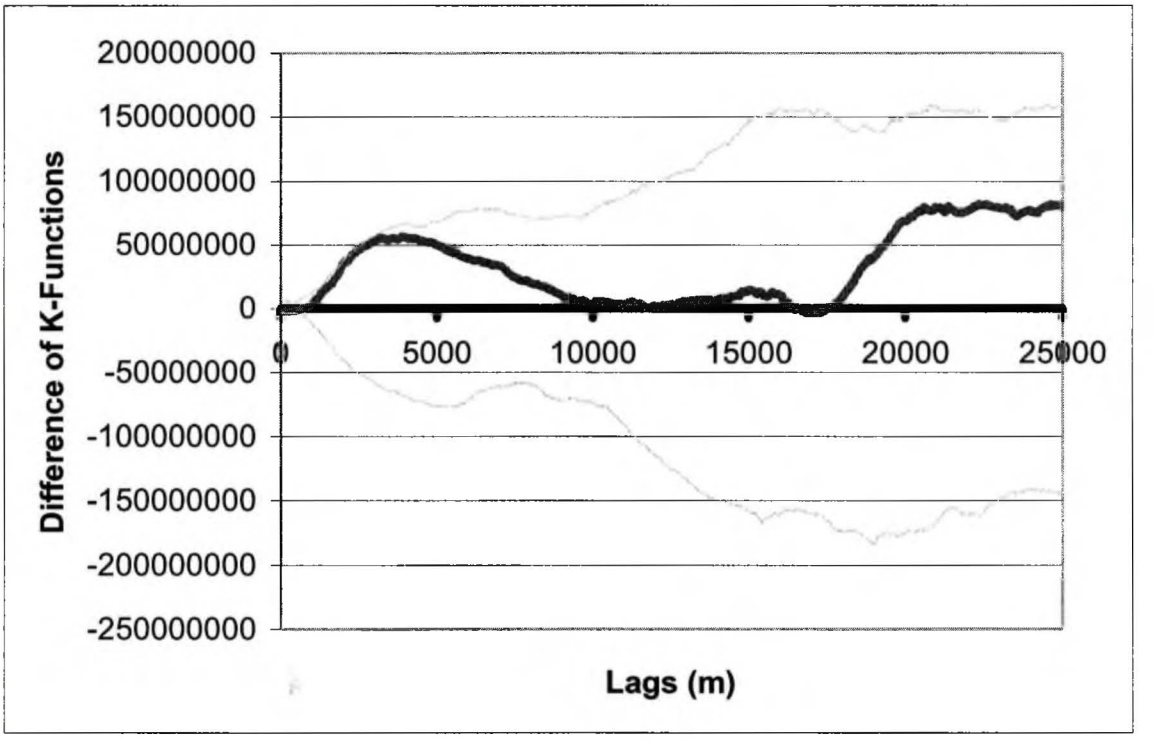


Figure 4.6. Difference of K-functions for preterm births, 1988-2002.

In figure 4.6 it is evident that the observed statistic (represented by the thick line) lies between the upper and lower simulation envelopes at all lags, (represented by the thin lines), although it closely follows the upper simulation envelope from 0 to approximately 4 Km. This suggests that the two sets of events, preterm births and controls, could have come from the same population; meaning that there is no clustering of preterm births once the underlying population has been controlled for. In other words, the difference in clustering between the cases and the controls are random rather than systematic. The graphs for congenital anomalies, LBW and stillbirths all show the same

result, meaning that in each case, there is no significant difference between K-function for the cases and the K-function for the controls for the full time period (see Appendix IV for the graphs).

Each of the outcomes (except stillbirths, due to the small number of cases) were examined in five year intervals as well. This was done in the event that a signal may have been present in one of the five-year periods, which may have been negated in the later years. This is especially relevant in the case of Sydney since the primary source of exposure, the Coke Ovens, ceased operation in 1988. The first interval spans the years 1988-1992; the second covers the years 1993-1997 and the final interval covers 1998-2002. Figure 4.7 shows the result for the first time period for preterm births (n=597). The pattern of the observed is similar to that for the entire period, however, in this case, the observed is greater than the simulated result from 0 to 5 Km. At these short distances, there is a significant difference between the observed and the expected signifying clustering in preterm births beyond the background population. There is also significant clustering in the second time period (1993-1997, n=519) at two lags, approximately 10 km from an index case, and 20 km from an index case. This can be seen in figure 4.8.

The remaining plot for preterm births (1988-2002, n=488) has a striking feature that is different from all the others, that is the presence of consistently negative values (Figure 4.9). Diggle and Morris (1996) describe this as 'anti-clustering', but point out that due to the strong presence of spatial autocorrelation, the trend may be misleading. The Monte Carlo tests of significance make allowances for autocorrelation, and the empirical statistics was not statistically significant at any spatial scale.

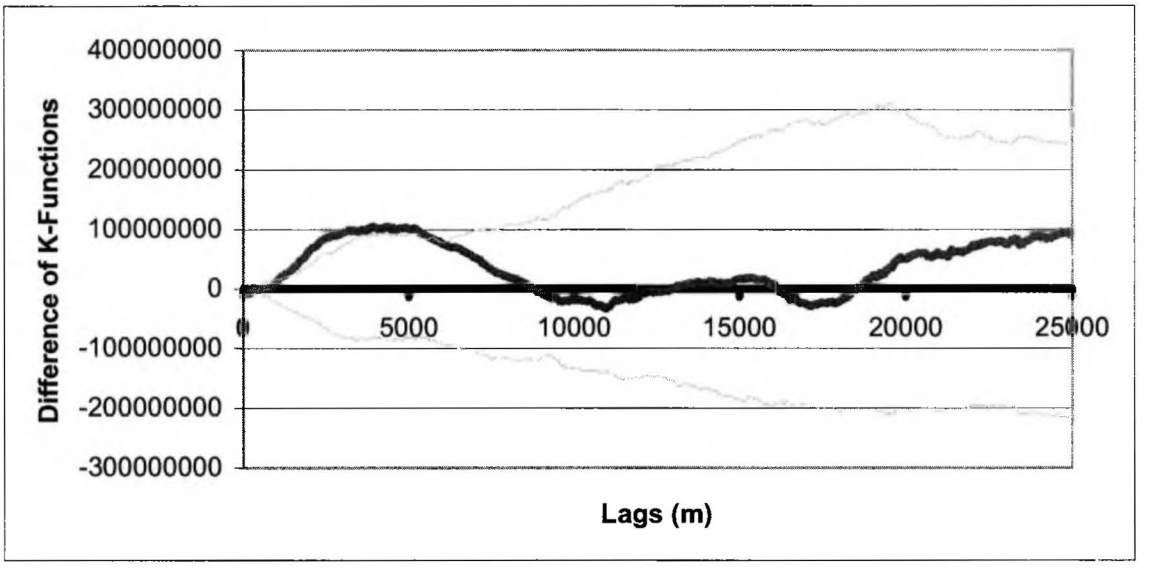


Figure 4.7. Difference of K-Functions for preterm births, 1988-1992.

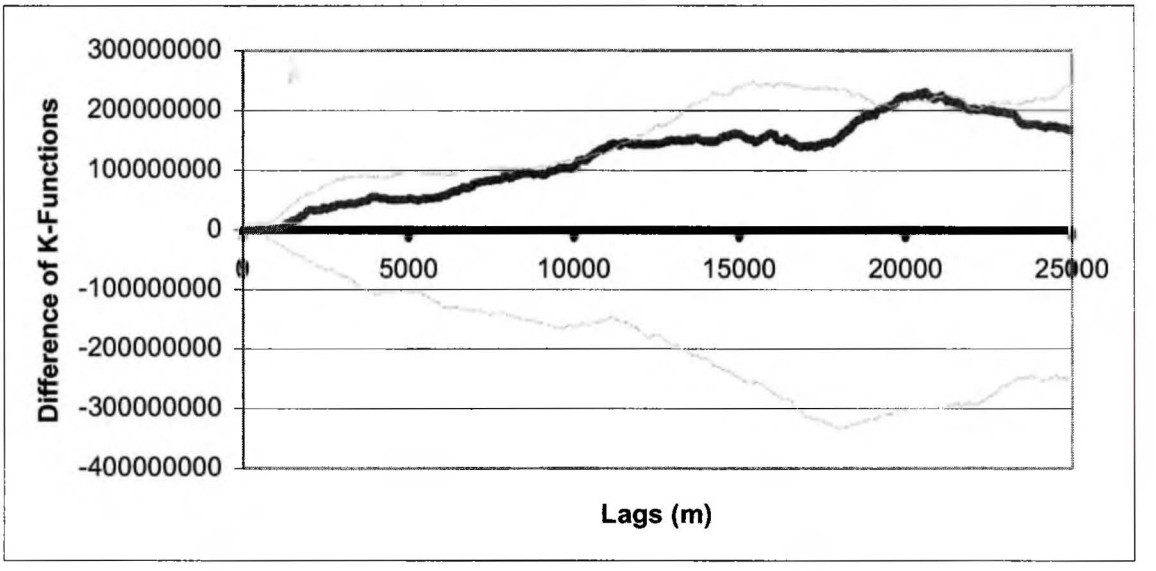


Figure 4.8. Difference of K-Functions for preterm births, 1993-1997.

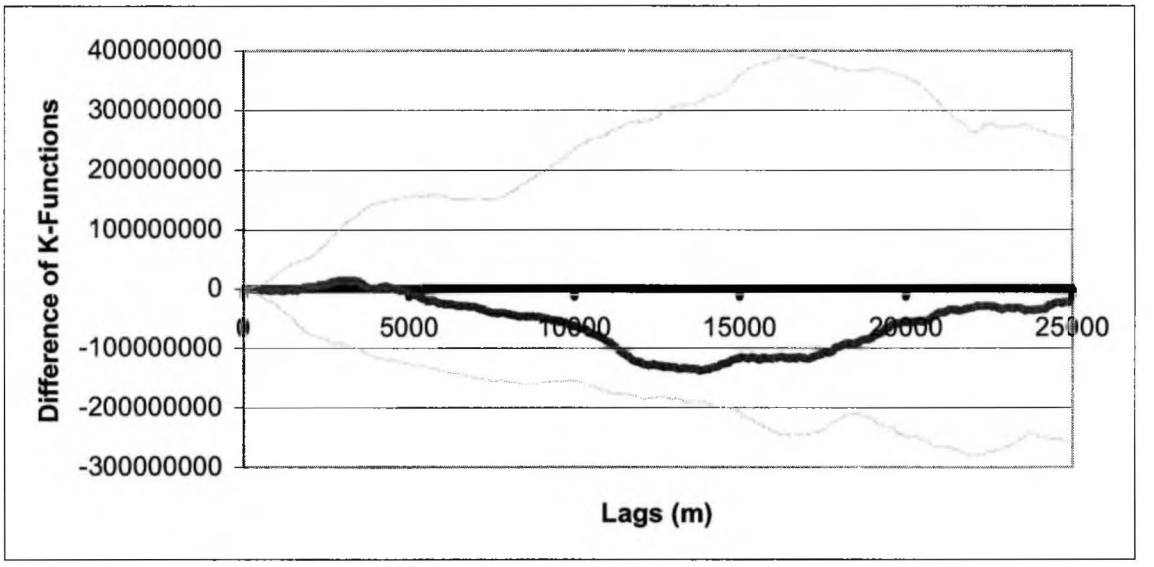


Figure 4.9. Difference of K-Functions for preterm births, 1998-2002.

The plots for the other outcomes for the five-year time periods failed to show any significant differences between the cases and the controls. These can be found in Appendix IV.

4.3.2 Ratio of Kernel Estimates

While the difference of K-Functions was a useful test to assess the overall spatial pattern of negative reproductive outcomes in CBRM, it does not identify where, if any, individual clusters occur. To determine where clusters might be located, the variation in the mean of the process was examined using a ratio of kernel estimates. A 1 km bandwidth was applied to each of the outcomes (in both the numerator and the denominator) and the results were mapped. Figure 4.10 shows the ratio of kernel estimates for preterm births for the years 1988-2002. The ratio of the two kernels is an estimate that measures the number of events per population, rather than events per area.

In this case, all kernel estimates are mapped as events per 1000 live births. In figure 4.10, all values less than 10 preterm births per 1000 live births were removed (see Chapter 3) leaving 812 grid cells upon which estimates were formed.

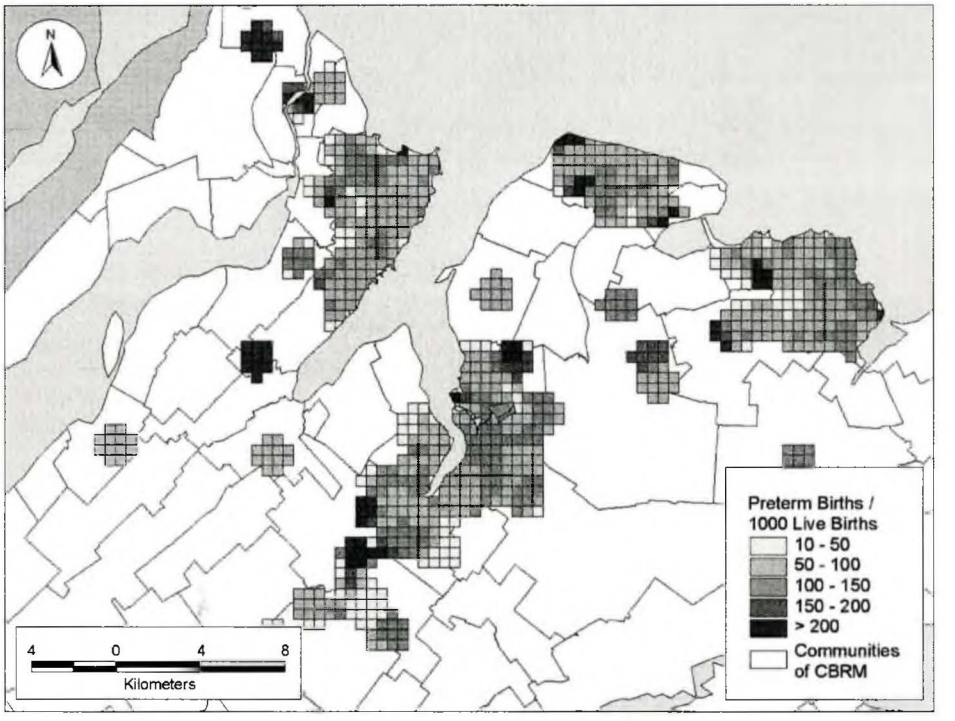


Figure 4.10. Ratio of kernel estimates for preterm births, 1988-2002.

It is clear that the mean of the process varies across space, with the majority of estimates between 50 and 150 preterm births per 1000 live births. The pockets of estimates that are greater than 200 preterm births are all areas where there was more than one event geocoded to an identical location, divided by very few events in the denominator. One interesting observation is that even though the underlying population has been accounted for, much of the area encompassing ICB exhibits elevated levels of preterm births.

To determine if the result in figure 4.10 is significantly different from random, 99 simulations were computed under the random labeling hypothesis (see Chapter 3). Figure 4.11 shows the result of the simulation process. After simulation, the pattern remains similar to that in figure 4.10, however, 231 grid cells were found to be insignificant and were removed. Significance was determined by comparing the observed value of the grid cell to the mean value of the corresponding grid cell generated through 99 simulations. Each of the grid cells in figure 4.11 satisfied the condition that the observed values was greater than two standard deviations away from the mean value of the 99 simulations. The areas closest to the Tar Ponds display levels of 50 to 150 preterm births per 1000. This pattern, while different than the results for Cape Breton County, is similar to the other areas of ICB, both in coverage (i.e., most of the respective administrative areas are covered by elevated estimates) and in magnitude (mostly 50 to 150 events per live births).

Figures 4.12 and 4.13 display the results for congenital anomalies. The pattern is similar to that of preterm births, where much of ICB displays elevated levels of anomalies per 1000 births. The ratio of kernel estimates for anomalies produced 699 grid cells after the filters described in chapter 3 were applied. After simulation, 467 grid cells remained (figure 4.13). There is variation present within the boundary of Sydney that differs from the result for preterm births. In figure 4.11, there is not a clear pattern of preterm births in that there are groups of cells in the 50 – 100 classification and the 100-150 classification mixed throughout Sydney. In figure 4.13, there appears to be a grouping of cells in the 100-150 classification north of the Tar Ponds in Whitney Pier.

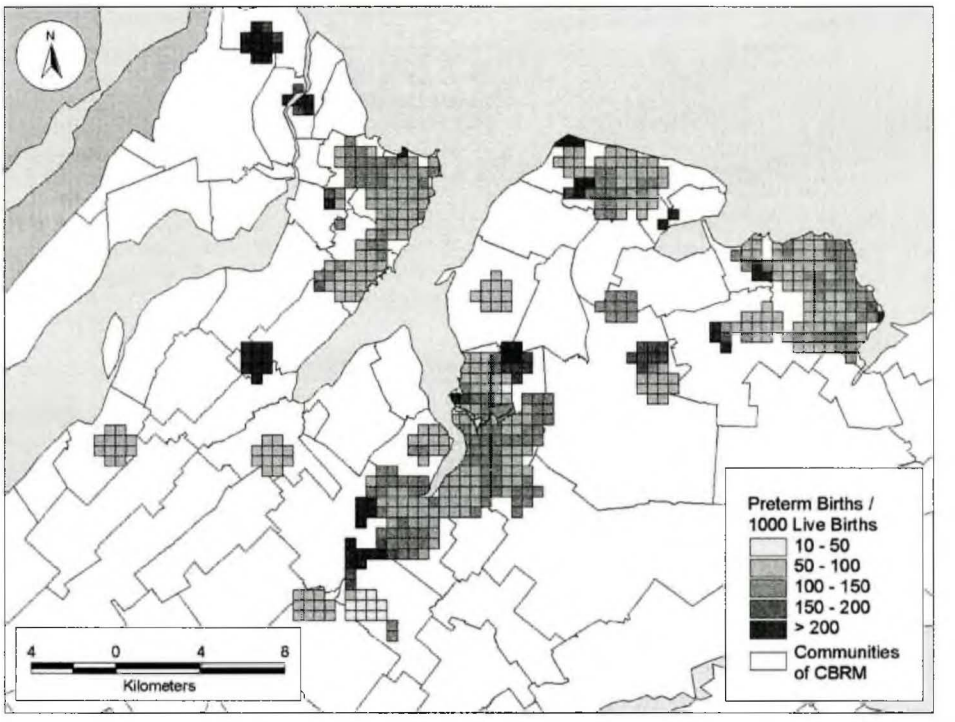


Figure 4.11. Statistically significant grid cells as determined through 99 simulations for preterm births

These elevated levels are consistent throughout much of the neighbourhood. To the south of the Tar Ponds in Ashby, the estimates are consistently lower (ranging from 50-100 anomalies per 1000 births). Further south, in Hardwood Hill and South End (southern part of Sydney) the levels are elevated again (100-150 anomalies per 1000). In the other communities of ICB, the predominant level of anomalies per 1000 births is between 50 and 150.

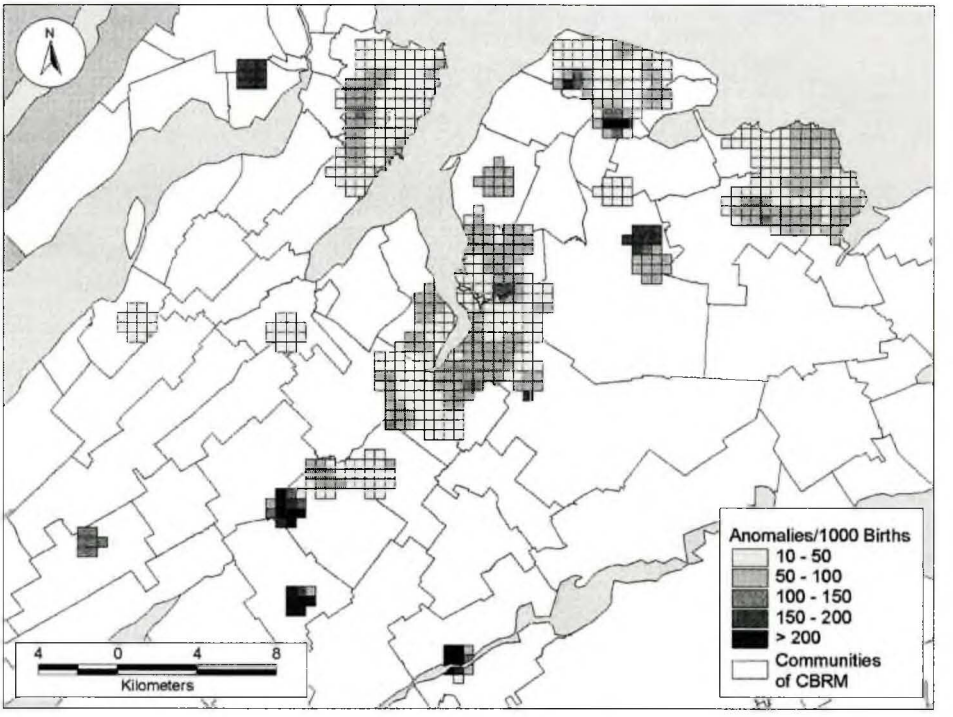


Figure 4.12. Ratio of kernel estimates for congenital anomalies, 1988-2002.

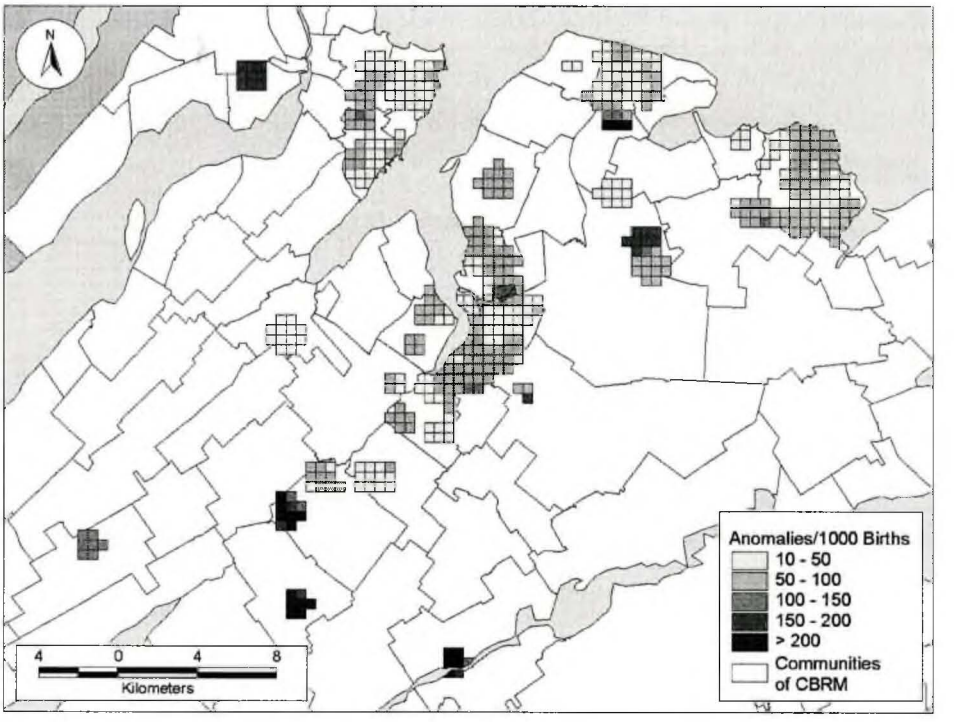


Figure 4.13. Statistically significant grid cells as determined through 99 simulations for congenital anomalies.

The results for LBW and stillbirths are quite different from the previous two outcomes both in coverage and in magnitude. This is due to the smaller numbers of events in the numerator (n=185 and n=100 respectively). The ratio of kernels for LBW (figure 4.14) produced 316 grid cells after the filters were applied. Approximately half of the grid cells were found to be insignificant through simulation and were removed. This result can be seen in figure 4.15. There are marked differences in the patterns displayed here and for the previous two outcomes. In this case, there are very few estimates formed in Sydney Mines and North Sydney (to the northwest). However, significant clusters can be seen downwind from the Coke Ovens site (northeast) in New Waterford and Glace Bay. The magnitude of the estimates have decreased in comparison to the previous two outcomes with the mode equal to 10-20 events per 1000 live births. Within Sydney, much of Whitney Pier does not have an estimate of LBW per 1000 live births. North End, Hardwood Hill and parts of Ashby do experience some elevated levels of LBW.

Stillbirths are relatively rare phenomena in most western societies, so it is not surprising that it has the lowest estimates per 1000 births. The mode of this outcome is the same as for LBW (10 – 20 per 1000 births). Figure 4.16 displays the results for stillbirth after simulation. Prior to simulation, 169 estimates were retained after the filters were applied. After simulation this was reduced to 161 estimates. Since there was very little difference between the two results, only the results after simulation are presented. Within Sydney, there is a pocket of estimates in Whitney Pier, but the majority of estimates occur south of the Tar Ponds in Hardwood Hill. In the other communities of ICB, the coverage is not as great as several areas do not have any estimates.

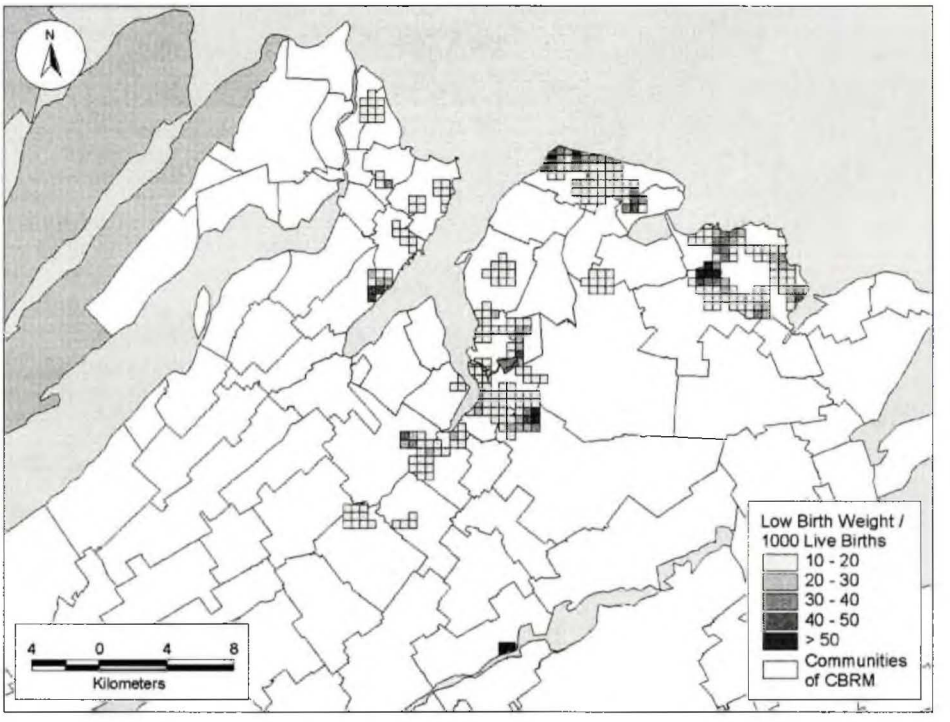


Figure 4.14. Ratio of kernel estimates for LBW, 1988-2002

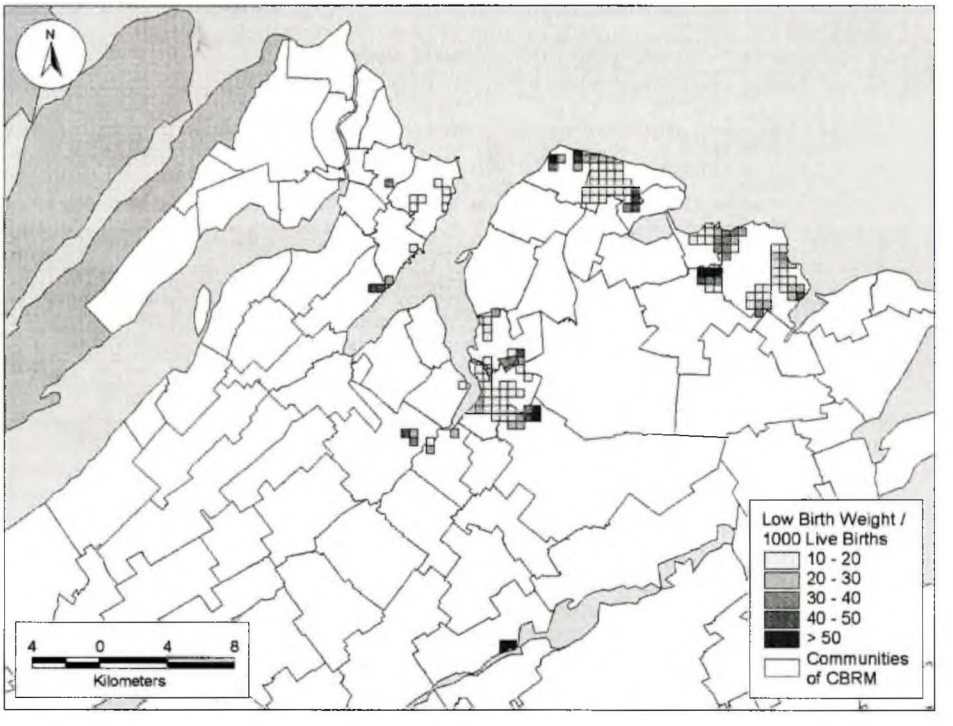


Figure 4.15. Statistically significant grid cells as determined through 99 simulations for LBW

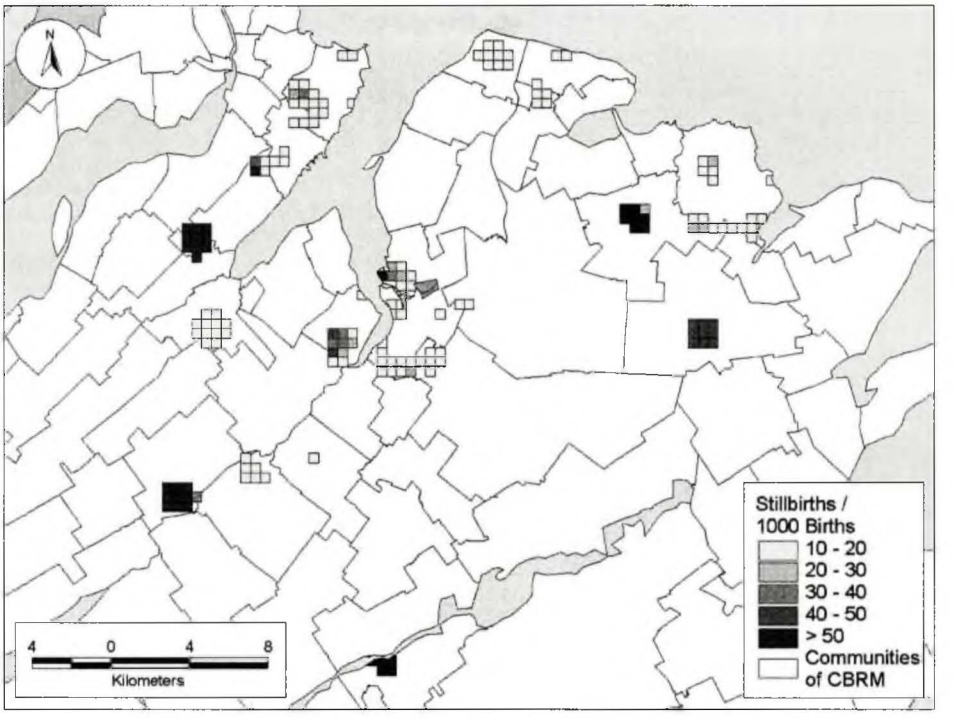


Figure 4.16. Statistically significant grid cells as determined through 99 simulations for stillbirths.

4.4 Discussion and Summary

Section 4.2 of this chapter presented the results of the geocoding and visualization applied to the data to help address the first research objective:

i. To assess the spatial pattern of various types of adverse reproductive events, plausibly linked to the environmental exposure of interest;

As described in Sections 3.4 and 4.2, a geocoding algorithm was applied to improve the resolution of the geocodes originally supplied with the data set. 17% of all records were correctly matched to an address from the civic address point file provided by CBRM, resulting in the best possible resolution. The bulk of the data set (80% of all records) was re-geocoded using a PCCF from DMTI Spatial. This value-added PCCF provided a marked improvement over Statistics Canada's PCCF originally used. Records

geocoded in this fashion were now matched to the center of the postal walk rather than the most representative block face within an enumeration area. Only 3% of all records retained the original coordinates assigned to them through the algorithm applied at IWK Grace Hospital (see Section 3.4). The improvement in the geocoding is most noticeable in the visualization stage (Section 4.2.2), but is critical for all stages of the analysis as illustrated by Burra *et al.* (2002).

Although there was improvement in the positional information, there were still errors present in the geocoding that may have had an impact on the results. While Dodds and Seviour (2001) highlight that the Atlee Database is of good quality and contains reliable data, there are still errors that cannot be controlled for. For example, as highlighted in table 3.1, some of the methods used to geocode could not be improved upon. Any record with a value of 0, 7,8,9 or 10 lacked sufficient information for the geocoding process. These codes corresponded to records with missing postal codes or missing addresses. At best, these records had a place name such as Marion Bridge or Point Aconi and were geocoded to their respective place name with no possibility for improvement. These types of records are included in the 539 records where original coordinates were retained and represent less than 3% of the remaining records (see table 4.1). Secondly, any postal codes classed as rural (coded with an 11 in table 3.1) have little room for improvement by geocoding with a PCCF. Records with this type of postal code were all geocoded to a unique location per postal code. A good example of this is Marion Bridge (directly south of Sydney; the area can be seen in Figure 4.16 as a heavily clustered area toward the bottom of the map) where approximately 450 records are

geocoded to the same location since it is serviced by a postal outlet. Unless more detailed address information was provided no improvement could be achieved.

The majority of records were either address matched or geocoded using the PCCF from DMTI Spatial. Due to temporal constraints during the data acquisition, any records that had detailed address information but did not match the civic address point file (e.g. spelling mistakes, incorrect road type etc.) were geocoded with the PCCF. This could be improved upon by going through the database record-by-record and correcting the various mistakes. The improvement offered however, is uncertain given the spatial scale of the analysis and the fact that the spatial referencing offered from DMTI is correct within 50 meters in urban areas, and is a subject of further research.

Section 4.3 describes the exploratory spatial analysis carried out to help address the first two objectives. The bivariate K-function proposed by Diggle and Chetwynd (1991) proved to be useful in detecting significant clustering at some spatial lags for preterm births in the first two temporal sub-sets. Results for the full time period demonstrated a weak tendency to cluster but failed to produce statistically significant results. Table 4.2 summarizes the results of the other outcomes including the temporal subsets. There are some instances where the empirical statistic straddles the line of no difference between the cases and the controls. At some spatial lags, there may be a weak indication of clustering, while at other lags there may be weak repulsion. This table presents the dominant result of each graph

Table 4.2. Results of the bivariate K-function for all outcomes.

Year	Outcome			
	Preterm	Anomalies	LBW	Stillbirths
1988-2002	Weak Clustering	Weak Clustering	Weak Clustering	Weak Clustering
1988-1992	Clustering*	Weak Repulsion	Weak Clustering	n/a
1993-1997	Clustering*	Weak Clustering	Weak Clustering	n/a
1998-2002	Weak Repulsion	Weak Clustering	Weak Repulsion	n/a

* indicates statistical significance at some spatial lags as determined through Monte Carlo simulation

The majority of the results indicate weak clustering (i.e., positive values of the test statistic, but insignificant). In three cases there was weak repulsion (preterm births and LBW for the 1998-2002 temporal subset and anomalies for the 1988-1992 subset). Two cases showed significant clustering at some spatial scales (preterm births for the first two temporal subsets). An analysis for stillbirths was conducted only for the full time period due to the small number of cases and failed to produce significant results.

The bivariate K-Function is used to address whether the cases are more clustered than the controls across CBRM. In this case, it is evident that only preterm births yield a significant result. Perhaps, not so coincidentally, the only reproductive outcome to yield any significant results in Burra (2002) was preterm births.

Dolk *et al.* (1998b) applied Diggle and Chetwynd's (1991) statistic to cases of anophthalmia and microphthalmia in England (1988-1994). They found that the test

failed to produce statistically significant results on a national level. They also performed the analysis on a subset of data, but in this case the subsets were defined spatially rather than temporally (the two however are analogous). They found the statistic exhibited clusters within 2 Km of an index case in Trent and within 50 Km of an index case in Oxford. This finding parallels the results in this study where the analysis produced muted results for the full data set but yielded some significant findings (for preterm births) in two different time periods.

Apart from the problems in geocoding mentioned previously, a source of error in this method may have been in the selection of controls. Cuzick (1998) states the importance of properly accounting for the variation in the underlying population. He suggests that the complete population can be used when available but when the scale is very small it is more accurate to use a sampling scheme for selecting controls. In this case, the scale of analysis is relatively large, so as a test, the statistic was calculated for congenital anomalies using the entire collection of normal births as the control population. The result of this test (after 24 hours of calculations) was virtually identical to the statistic calculated using the controls selected in the manner described in chapter 3 and advocated by Gatrell (2002).

A second issue with this method raised by Cuzick (1998) is how to define closeness. Diggle and Chetwynd (1991) extend the K-function to deal with a bivariate point pattern. The root of their statistic is the K-function, where discrete bands of fixed distances were created, and events within each distance band were counted. Reader (2000) points out that this is a form of aggregation and data loss, since it does not matter

where within the distance band an arbitrary event lies in relation to the index case. Although this is unavoidable, it is currently the best practice for point pattern analysis since it measures spatial dependency on a wide range of scales (Reader, 2000).

The results described in section 4.3.2 dealt with the next logical step in the analysis; clustering on a local scale. Although the analysis on a global scale produced very few significant results, it was still useful to examine clustering on a local level. This is analogous to the situation where a global indicator of spatial autocorrelation, such as Moran's I, fails to produce a significant result, but the local form of the statistic yields significant areas within the dataset.

The results presented in section 4.3.2 build off the work of Reader (2001) and Rushton and Lolonis (1996) where a grid was used to generate estimates of negative outcomes that vary continuously through space. A further similarity is the use of the random labeling hypothesis within a Monte Carlo simulation framework for inference across the three studies. Similar to Reader (2001), kernels rather than simple rate calculations (as in Rushton and Lolonis, 1996) provided the mode of estimation. Unlike Reader (2001) a ratio of kernel estimates was applied to the data in an attempt to correct for the underlying population density. Reader (2001) was interested in identifying clusters of high rates of incidence of LBW. Under the random labeling hypothesis, he applied a basic kernel estimate with a 4 km bandwidth and found significant clusters in several parts of central Florida and the Miami area. Although the Monte Carlo techniques to determine significance are an improvement over the basic kernel estimate, it still provides the number of events per unit area. This still suffers from the fact that each point

pattern randomly selected for the simulations are driven by the underlying density of births. The results produced in this research on the other hand, produce an estimate of the number of events per 1000 births, thereby correcting for the underlying population.

When comparing the results to the provincial and national rates, some interesting pictures appear. Table 4.3 presents the rates of various reproductive outcomes per 1000 for Nova Scotia and for Canada (modified after Burra, 2002). Table 4.4 presents the prevalence rates of the outcomes for the 15 year time period, as calculated from the Atlee Perinatal database for the areas of ICB plus Louisbourg and the rest of Cape Breton County. As in Burra (2002), the time lines do not perfectly match, however, some interesting findings still emerge. Each of the figures depicting results for the ratios of kernel estimates demonstrate local pockets that have reliable estimates that are higher than both the provincial and national rate. In particular, for congenital anomalies, the provincial rate is 23 per 1000 births, the prevalence for Sydney is 49.23 per 1000 (from Table 4.4), while the mode of the kernel estimates within Sydney's boundaries (figure 4.13) is 50-100 events per 1000. For preterm births, the provincial rate is 58 per 1000 live births, while the prevalence from the Atlee database over the 15 year time period is 94.87/1000 for Sydney. Figure 4.11 shows several estimates within 50-100 events per 1000 as well as several estimates above the provincial rate in the 100-150 events per 1000 range. The results of the kernel ratios for Glace Bay produce similar estimates to those for Sydney for all outcomes, while the other parts of ICB are generally lower, and the remainder of CBRM has minimal grid cell estimates and lower prevalence rates than Sydney for all of the outcomes (Table 4.4).

Table 4.3. Prevalence of pregnancy outcomes in Nova Scotia and Canada (after Burra, 2002)

Pregnancy Outcome	Nova Scotia Rate/1000	Canada Rate/1000	Source(s)
Preterm birth	58 (1988-98)	70 (1991-2000)*	Dodds and Seviour, 2001 Health Canada, 2000
Congenital Anomalies	23 (1988-98)	Not Available	Dodds and Seviour, 2001
Low Birth Weight	54 (1979-99)	56 (1987-94)	Joseph and Kramer, 1997 NSRCP, 2000
Stillbirth	2.6 (1988-98)	4.6 (1985-99)*	Health Canada, 2000 NSRCP, 2000

* excluding data for Newfoundland and Ontario due to missing data and data quality issues respectively

Table 4.4 Prevalence rates per 1000 of reproductive outcomes as calculated from the Atlee Perinatal Database.

Location	Preterm Birth	Congenital Anomalies	Low Birth Weight	Stillbirth
Sydney	94.87	49.23	11.20	7.10
Sydney Mines	90.73	39.90	7.78	7.08
North Sydney	77.53	45.89	6.10	6.06
New Waterford	92.23	47.46	14.52	7.63
Dominion	95.96	30.30	15.15	0.00
Glace Bay	101.06	52.59	13.03	4.21
Louisbourg	86.46	43.23	5.76	0.00
Cape Breton County (remainder of County)	87.00	46.98	9.65	5.22
TOTAL	**91.41**	**47.31**	**10.54**	**5.67**

The selection of the bandwidths for the numerator and the denominator were based on experimentation. A bandwidth of 1 km was decided upon for both the numerator and the denominator. Gatrell *et al.* (1996) suggest that it is not necessary to have the same bandwidth for the numerator and the denominator. They even suggest increasing the bandwidth of the denominator several fold over the numerator in order to adequately smooth out the estimate. This is preferable in instances where proxies for the underlying population are used as in Buliung and DeLuca (2000). Since this research was based on a marked point pattern representing the full enumeration of events, this was not necessary. The ramifications of using a fixed bandwidth of 1 km are that some of the grid cells in rural areas outside of ICB need to be interpreted with caution. In some instances, there was an isolated event, and that event happened to be negative. The resulting ratio would be unity, affecting not only the grid cell closest to the event but also all adjacent grid cells since a weighting scheme is applied to all events within 1 km. Cases such as these, would not be corrected for in the simulation process, since the odds of that event being selected enough times to eventually render it insignificant are virtually nil. Adaptive bandwidths were also attempted as a way to address the spurious clusters in the rural areas and on the perimeter of the urban areas. These were attempted in the exploratory phases of this research, but not applied since some of the outcomes had small numbers for the numerator. Additionally, the errors in geocoding would have in some cases yielded even more unreasonable estimates.

The spatial component of this research was useful in that on one hand, it helped to confirm the findings of Dodds and Seviour (2001), where they found a slight increase in

congenital anomalies in Sydney versus the region. On the other hand, it extended their work by pinpointing areas within not only Sydney, but the remainder of ICB that are elevated over the background rate, not only for congenital anomalies, but for all of the outcomes under study. Analyzing the data using these methods are powerful since they do not rely on administrative units as most traditional epidemiological studies do (Rushton, 2003). In fact Dunn *et al.* (2003) point out that the types of methods applied in this research are complementary to epidemiological studies and advocate the use of point pattern analysis in studies dealing with spatially referenced public health data.

The methods applied in this chapter were chosen to address the first research objective. In addressing the first objective, the results of the bivariate K-function demonstrate that for preterm births, there is clustering at some spatial lags on a global scale. The kernels indicate that the spatial pattern of adverse reproductive outcomes is not random, especially for preterm births and congenital anomalies since all of the Sydney and the majority of ICB have elevated estimates of these outcomes per 1000 births. The results for LBW and stillbirth are also non-random however the pattern is not as dominant as it is for preterm births and congenital anomalies.

The second objective was to determine if the spatial pattern is related to proximity to the Tar Ponds/Coke Ovens Site. There is not enough evidence to suggest this from a purely spatial perspective, since the areas closest to the Tar Ponds show similar patterns to those quite some distance away. There is nothing unique about the intensity of the events around the Tar Ponds.

In order to better address the second objective, multivariate methods were applied to the data. One of the hypotheses to be tested is the significance of proximity to the Tar Ponds. Proximity is measured in two ways, first through zones around the point source as in Bhopal *et al.* (1999) and second, measuring proximity through a continuous variable from a reproductive event to the Tar Ponds. In addition to distance, several other covariates will be tested in a MNL model in order to control for known factors that have been previously associated with any of the adverse outcomes under study.

CHAPTER FIVE

MULTIVARIATE ANALYSIS OF ADVERSE REPRODUCTIVE OUTCOMES

5.1 Introduction

This chapter presents the results of the multivariate analysis of the Atlee Perinatal Database. In the previous chapter, exploratory spatial data analysis was employed to address the first research objective. The spatial measures, however, were not sufficient to address the second question and the multivariate methods in this chapter may prove helpful in that regard. With that in mind, the methods of this chapter have been used to address the second research objective:

ii. To determine if the pattern of adverse outcomes is related to proximity to the Tar Ponds/Coke Ovens site

With respect to this objective, bivariate methods incorporating various indicators of proximity to the Tar Ponds were tested to determine if the patterns observed in chapter four are related to proximity to the Tar Ponds. Additional covariates were included in the modeling to control for factors such as smoking, marital status, pregnancy history and other pregnancy-related measures (see Appendix V for variable definitions).

5.2 Descriptive Analyses of Adverse Reproductive Outcomes

Table 4.4 in the previous chapter presented the prevalence rates for each of the outcomes under study for the entire study period (1988-2002). The prevalence of preterm births in Sydney was about 9.5% of all live births during that time. By comparison, Dodds and Seviour (2001) found the prevalence of preterm births to be 6.6% for 1988-1998 within Sydney. The increase observed with the addition of the years 1999-2002 is

consistent with the findings of Health Canada (2003), where preterm births continued to steadily rise with each passing year. The Canadian prevalence rate of preterm births in 1991 was 6.6% and continued to rise each year reaching 7.6% by 2000. When comparing Sydney to Canada and to Nova Scotia (Table 4.3), it can be seen that the prevalence is higher in Sydney. Interestingly, the remainder of ICB and CBRM show higher prevalence rates of preterm birth than the province and Canada also, with Glace Bay having the highest at approximately 10%. In another study, Yang *et al.,* (2002) found the prevalence of preterm birth in areas exposed to petrochemical industries to be significantly higher ($p < 0.05$) than those in non-exposed areas (5.13% versus 4.46%).

Approximately 5% of all births in Sydney during the study period were associated with congenital anomalies. This was greater than the figures reported by Dodds and Seviour (2001) for Nova Scotia (2.3%), Sydney (2.8%) and Cape Breton County excluding Sydney (2.3%). It is also higher than the expected frequency of 2-3% of all births described by Bloom (1981) for studies of populations exposed to reproductive hazards. In fact, all places in ICB except for Dominion have higher prevalence rates of congenital anomalies. These elevated figures may be explained by the manner in which births were coded as congenital anomalies (see chapter 3). For this study, all anomalies were grouped together into one variable whereas Dodds and Seviour (2001) have examined major anomalies and several sub-categories of anomalies. In other studies, Bhopal et al (1999) found the prevalence of major congenital anomalies to be 1.6% in close proximity to steel industries in Teeside, UK, while Elliott et al (2001) found the prevalence of all congenital anomalies to be 1.7% in close proximity to landfill sites.

Fielder et al (2000) on the other hand found the prevalence of all reported congenital anomalies to range between 1 and 6% for women living in areas with landfills accepting hazardous materials.

The prevalence of LBW babies was considerably lower for Sydney in this data set (about 1.05%) compared to the 5.6% reported by Dodds and Seviour (2001). In addition, the observed value in this study is below the values indicated by Bloom (1981) for populations exposed to reproductive hazards (7% of all live births). The reason for this difference is explained by the coding of the outcomes (Chapter 3). The prevalence of LBW in this case is controlled for congenital anomalies, preterm births and stillbirths (i.e., they were forced into mutually exclusive categories, so if a newborn was preterm and LBW, it was coded as preterm). This is not the case in Dodds and Seviour (2001) since they control for major anomalies, but not gestational age. In other words, Dodds and Seviour (2001) make no distinction between preterm LBW and term LBW as in this study. In other studies, Tough et al (2001) found the prevalence of LBW to be 6.4% of all live births in Alberta. Elliott et al (2001) found the prevalence of LBW to be 6.4 for women living near landfills, while Bhopal et al (1999) found the prevalence of LBW to increase as one moved from the most proximate zone to industry (7.8%) to the most distant zone (8.7%).

The rate of stillbirth in Sydney was 0.71% for the period 1988-2002. This is well below the rate of two to four per cent of all births described by Bloom (1981) for populations exposed to reproductive hazards. The remainder of ICB is also well below this rate (see Table 4.4). However, when compared to the rates published by Health

Canada (2003), it is evident that Sydney is more than double the Provincial rate of 0.26% and also greater than the National rate (0.46%). Interestingly, Glace Bay, which is amongst the highest in terms of prevalence for all other categories in ICB, is amongst the lowest for stillbirths at 0.42%. Bhopal et al (1999) found the prevalence of stillbirths in areas close to steel mills to be 0.5%. Elliott et al (2001) also found a similar prevalence of stillbirths that occurred in close proximity to landfill sites in the UK (0.51%).

5.3 Results of Bivariate Analysis

To help address objective ii, contingency analysis was conducted with each of the outcomes under study and various measures of proximity to the Coke Ovens site. This was done to determine if there was a higher risk of adverse reproductive outcome associated with more proximal distances to the point source of pollution. A GIS was used to create a buffer zone consisting of three concentric rings of a one-kilometer radius each around the point source as in Bhopal et al (1999) (i.e., from 0 to 1 km from the coke ovens, from 1 – 2 km and from 2 – 3 km). In addition, observations were grouped geographically into three zones for the purposes of comparison; Sydney, all of ICB excluding Sydney, and the remainder of CBRM. This was done in a similar fashion to Guernsey et al (2000). The results of this analysis are presented in Table 5.1. For the distance-based zones, there was a clear gradient of decreasing ORs with increasing distance away from the Coke Ovens site demonstrated only for stillbirths, however, the ORs were not statistically significant at any of the distances due to small numbers of events occurring within each zone. For preterm births, where the greatest odds occurred in the zone 2-3 kilometers away from the Coke Ovens site (OR = 1.23 (95% CI: 1.02 –

Table 5.1. Odds ratios derived from contingency analyses of adverse reproductive outcomes and proximity to the Coke Ovens site.

	DISTANCE-BASED MEASURES				COMMUNITY-BASED		
OUTCOME	**0 - 1 Km** (n=399)	**1 - 2 Km** (n=1779)	**2 - 3 Km** (n=1327)	**0 - 3 Km** (n=3505)	**Sydney** (n = 3902)	**ICB minus Sydney** (n = 6187)	**Rest of CBRM** (n = 5711)
Preterm Birth							
% Preterm	9.0	9.9	10.6	10.0	9.50	9.10	8.20
p-value*	0.93	0.12	0.03	0.01	0.11	0.46	0.03
Odds Ratio (95% CI)	1.02 (0.72 - 1.44)	1.14 (0.97 - 1.35)	1.23 (1.02 - 1.48)	1.19 (1.05 - 1.35)	1.11 (0.98 - 1.26)	1.04 (0.93 - 1.17)	0.88 (0.78 - 0.99)
Number of Cases	36	176	140	352	372	563	470
Congenital Anomaly							
% Anomaly	6.3	4.2	5.1	4.8	5.2	4.8	4.6
p-value*	0.17	0.17	0.59	0.86	0.23	0.86	0.38
Odds Ratio (95% CI)	1.33 (0.88 - 2.01)	0.84 (0.66 - 1.08)	1.07 (0.83 - 1.39)	0.98 (0.83 - 1.17)	1.11 (0.94 - 1.30)	0.99 (0.85 - 1.15)	0.93 (0.80 - 1.09)
Number of Cases	25	74	68	167	202	296	264
Low Birthweight							
% < 2500g	1.3	1.0	1.3	1.1	1.1	1.1	0.9
p-value*	0.67	0.95	0.33	0.44	0.47	0.46	0.16
Odds Ratio (95% CI)	1.23 (0.50 - 3.02)	0.99 (0.60 - 1.61)	1.28 (0.77 - 2.13)	1.15 (0.80 - 1.65)	1.14 (0.80 - 1.61)	1.13 (0.82 - 1.54)	0.79 (0.56 - 1.10)
Number of Cases	5	18	17	40	44	68	50
Stillbirth							
% stillborn	1.0	0.7	0.2	0.5	0.6	0.5	0.4
p-value*	0.16	0.29	0.13	0.74	0.27	0.88	0.25
Odds Ratio (95% CI)	2.04 (0.74 - 5.61)	1.39 (0.75 - 2.58)	0.42 (0.13 - 1.34)	1.09 (0.65 - 1.83)	1.31 (0.81 - 2.11)	1.04 (0.66 - 1.62)	0.76 (0.47 - 1.22)
Number of Cases	4	12	3	19	24	32	24

* p-value determined from a Chi-squared test

1.48)). This was the only statistically significant relationship for any of the outcomes and the three concentric zones. The sum of the zones was also examined (i.e., from 0 – 3 kilometers) yielding a statistically significant relationship with preterm births (OR = 1.19 (95% CI: 1.05 – 1.35)).

The results of the contingency analysis for place of residence and an adverse outcome produced an interesting pattern. They show that Sydney always had the highest odds ratio for each outcome, followed by ICB excluding Sydney, and finally the remainder of CBRM had the lowest odds of having an adverse outcome (Table 5.1). However, the only statistically significant result from this cross-tabulation was for preterm births, where the odds of having an adverse event are significantly lower in the more rural areas of the county.

Proximity was also examined using the distance to the coke ovens site as a continuous variable as in Dolk et al (2000). Here, mean distances were compared using a t-test to examine significant differences between the mean distance of all cases and non-cases to the polluted area. Table 5.2 demonstrates that the only statistically significant difference between the two groups occurred for stillbirths, with the cases having a shorter mean distance to the coke ovens site than the non-cases.

Table 5.2. Comparison of mean distances of cases and non-cases to the coke ovens site.

Outcome	Mean of Non-Cases	Mean of Cases	t-test*	p-value
Preterm Birth	10.67	10.56	0.549	0.583
Anomalies	10.66	10.69	-0.106	0.916
LBW	10.66	10.68	-0.023	0.982
Stillbirth	10.67	9.27	2.048	0.044

* t-test assumed unequal variances

The same analysis was conducted with all cases using maternal first pregnancy only (as outlined in chapter 3). This, however, failed to produce any significant associations with any of the distance variables. See Table A6.1 in Appendix VI for the results of this analysis.

Contingency analysis was also performed on several variables obtained from the Atlee Perinatal Database or from the 1996 census. Table 5.3 presents the results of this analysis. Several significant associations were observed using a chi-squared test. For maternal characteristics, women who were legally married or common-law had significantly lower odds of having a preterm birth than those who were single, widowed, divorced or separated at the time of delivery. With regards to maternal age, women who were 18 years of age or younger had an increased odds of having a stillbirth (OR: 1.81 (95% CI 1.03 – 3.19). Age (either 18 and younger or 35 and older) failed to produce a significant relationship for any of the other age categories. Smoking during pregnancy produced significant increases in odds of having a preterm birth, LBW and stillbirth with the largest odds occurring for LBW (OR: 2.67 (95%CI 1.95-3.66)). Maternal prescription drug use also proved to be important, with significant ORs for preterm births, congenital anomalies and stillbirths. Maternal drug abuse was significant for preterm births only. Finally, residing in an enumeration area that was amongst the most deprived areas in CBRM proved to be insignificant for all of the outcomes under examination.

Pregnancy history has been shown in the literature to be a significant indicator of having an adverse reproductive outcome (ACOG, 2001, Kramer, 1987a, Kramer, 1987b,

Table 5.3. Odds ratios derived from contingency analyses of adverse reproductive outcomes with various covariates (n = 15800)

Covariate	n	Preterm Birth (n = 1405)		Anomaly (n = 762)		LBW (n = 162)		Stillbirth (n = 80)	
Married	9453	0.77 (0.69 - 0.86)	***	1.03 (0.88 - 1.20)		0.62 (0.46 - 0.85)	**	1.01 (0.64 - 1.58)	
Under 19 yrs of Age	1789	1.13 (0.95 - 1.33)		0.97 (0.77 - 1.22)		1.30 (0.83 - 2.03)		1.81 (1.03 - 3.19)	*
Over 34 Yrs of Age	1387	1.11 (0.92 - 1.33)		1.09 (0.85 - 1.40)		1.30 (0.80 - 2.13)		1.66 (0.88 - 3.15)	
5th Quintile of Deprivation Index	4355	0.93 (0.82 - 1.05)		1.04 (0.88 - 1.22)		1.01 (0.72 - 1.43)		0.94 (0.57 - 1.54)	
Ever Smoked During Pregnancy	5612	1.31 (1.17 - 1.47)	***	0.96 (0.82 - 1.12)		2.67 (1.95 - 3.66)	***	1.57 (1.01 - 2.43)	*
Maternal Drug Use	199	4.24 (3.11 - 5.80)	***	1.86 (1.13 - 3.08)	*	1.49 (0.47 - 4.70)		4.19 (1.52 - 11.57)	**
Maternal Drug Abuse	35	2.57 (1.12 - 5.89)	*	1.20 (0.29 - 5.00)		n/a		n/a	
Nulliparity or Grand Multiparity	7048	1.32 (1.19 - 1.48)	***	1.06 (0.91 - 1.22)		1.77 (1.29 - 2.43)	***	1.45 (0.93 - 2.25)	
Previous Abortion	3035	1.14 (0.99 - 1.30)		1.04 (0.87 - 1.25)		1.04 (0.70 - 1.53)		1.31 (0.78 - 2.20)	
Previous Stillborn	170	2.22 (1.49 - 3.31)	***	0.85 (0.34 - 1.81)		1.75 (0.55 - 5.53)		4.93 (1.78 - 13.64)	***
Previous LBW	442	3.11 (2.47 - 3.91)	***	1.29 (0.87 - 1.92)		2.57 (1.38 - 4.78)	**	3.37 (1.54 - 7.36)	***
Previous Negative	833	2.40 (1.99 - 2.89)	***	1.49 (1.12 - 1.97)	**	1.18 (0.62 - 2.25)		2.59 (1.33 - 5.04)	**
Previous Maternal Conditions	1312	1.11 (0.91 - 1.34)		1.22 (0.95 - 1.56)		0.65 (0.33 - 1.27)		1.23 (0.59 - 2.56)	
Admitted to Hospital with Complications	2941	1.17 (1.03 - 1.35)	*	0.98 (0.82 - 1.19)		1.04 (0.70 - 1.53)		1.09 (0.63 - 1.89)	
Current Chronic Hypertension	175	3.74 (2.66 - 5.27)	***	1.73 (1.00 - 3.00)	*	1.12 (0.28 - 4.54)		3.52 (1.10 - 11.27)	*
Pregnancy Induced Hypertension	1541	1.86 (1.60 - 2.17)	***	1.32 (1.06 - 1.65)	**	1.31 (0.86 - 2.09)		1.80 (0.99 - 3.27)	
Anemia	790	1.43 (1.15 - 1.78)	**	1.14 (0.83 - 1.57)		0.73 (0.32 - 1.65)		4.09 (2.28 - 7.31)	***
Other Current Conditions	430	2.58 (2.02 - 3.30)	***	1.79 (1.25 - 2.54)	**	9.04 (6.04 - 13.54)	***	1.40 (0.44 - 4.44)	
Placental Anomalies	860	1.10 (0.87 - 1.39)		0.75 (0.52 - 1.08)		0.67 (0.29 - 1.51)		1.41 (0.61 - 3.25)	
Fetal Clinical Wasting	312	4.09 (3.17 - 5.27)	***	3.28 (2.36 - 4.57)	***	19.21 (13.22 - 27.93)	***	0.63 (0.09 - 4.52)	
Infant of Diabetic Mother	699	2.45 (2.01 - 2.99)	***	1.59 (1.19 - 2.14)	**	1.12 (0.55 - 2.30)		0.27 (0.40 - 1.96)	

* $p < 0.05$ ** $p < 0.01$ ***$p < 0.001$

significance determined from Chi-squared statistic

Kramer 2003). In this analysis, several variables designed to capture various aspects of pregnancy history proved to be significant for each of the outcomes. First, nulliparity (i.e., never been pregnant before the current pregnancy) and grand multiparity (i.e., having more than 4 prior pregnancies) was significantly associated with both preterm births and LBW. Having a prior pregnancy result in a stillbirth was significantly associated with preterm births and stillbirths, with the odds of having a stillborn child almost 5 times higher when a prior pregnancy resulted in a stillbirth. A prior pregnancy that resulted in a LBW proved to be significantly associated with approximately a three-fold increase in the odds of preterm birth, LBW and stillbirth, while having any of the other adverse reproductive outcomes under study in the past was significantly associated with increased odds of preterm births, congenital anomalies and stillbirths. Having a prior abortion (induced or spontaneous) or having a prior medical condition such as hypertension, anemia, gestational diabetes or eclampsia failed to produce any significant relationships with any of the outcomes.

Variables for present pregnancy maternal diagnoses had significant relationships in several instances. For example, chronic hypertension was associated with almost a four-fold increase in the odds of having a preterm birth, while pregnancy-induced hypertension approximately doubled the odds of having a preterm birth. Significant associations were also found between chronic hypertension and congenital anomalies as well as stillbirths. Pregnancy-induced hypertension was also significantly associated with congenital anomalies. Anemia had a significant increase in the odds of preterm birth and produced a four-fold increase in the odds of having a stillbirth. Finally the sum

of all other causes of maternal pregnancy complications produced significant associations with preterm births (OR: 2.58), congenital anomalies (OR: 1.79) and LBW (OR: 9.04).

The final set of covariates dealt with neonatal measures. There were 312 instances of fetal clinical wasting (sometimes known as recent fetal starvation) in the dataset. This produced significant associations with preterm birth (OR: 4.09), congenital anomalies (OR: 3.28) and LBW (OR: 19.21). Lastly, being an infant of a diabetic mother resulted in significantly elevated ORs for preterm births and congenital anomalies.

The same analyses were carried out for the primiparous women (n = 6009) in this data set and produced a fewer number of significant associations (Table 5.4). For preterm births, the following variables led to statistically significant elevated ORs: smoking, maternal drug use, chronic hypertension, pregnancy-induced hypertension, all other maternal complications, fetal clinical wasting and infant of a diabetic mother. For congenital anomalies only three variables produced a statistically significant relationship. These are the sum of the maternal complications (not including anemia, chronic hypertension and pregnancy-induced hypertension), fetal clinical wasting and infant of a diabetic mother. For LBW, women aged 35 and older, smoking, the sum of maternal complications and fetal clinical wasting produced significant associations. None of the variables were statistically significant with stillbirths.

As a result of the contingency analyses, several significant bivariate associations were determined between several variables and the outcomes being studied. The next section of this chapter presents the results of the multivariate analyses, which combines the statistically significant variables from the bivariate analyses into a MNL model.

Table 5.4. Odds ratios derived from contingency analyses of adverse reproductive outcomes with various covariates among primiparous women (n = 6009)

Covariate	n	Preterm Birth (n = 589)	Anomaly (n = 290)	LBW (n = 78)	Stillbirth (n = 36)
Married	2691	0.95 (0.80 - 1.13)	1.08 (0.85 - 1.37)	0.81 (0.51 - 1.28)	0.99 (0.51 - 1.91)
Under 19 yrs of Age	1440	1.01 (0.83 - 1.23)	0.95 (0.72 - 1.26)	0.95 (0.56 - 1.62)	1.81 (0.91 - 3.56)
Over 34 Yrs of Age	208	1.09 (0.70 - 1.71)	1.33 (0.75 - 2.36)	2.36 (1.02 - 5.50) *	0.80 (0.11 - 5.84)
5th Quintile of Deprivation Index	1247	0.85 (0.69 - 1.07)	0.95 (0.71 - 1.28)	0.83 (0.47 - 1.49)	1.28 (0.60 - 2.72)
Ever Smoked During Pregnancy	1932	1.24 (1.04 - 1.48) *	0.95 (0.74 - 1.22)	1.82 (1.17 - 2.86) **	1.51 (0.78 - 2.94)
Maternal Drug Use	73	4.66 (2.84 - 7.65) ***	1.78 (0.77 - 4.14)	2.17 (0.52 - 9.02)	2.34 (0.32 - 17.33)
Admitted to Hospital with Complications	1169	1.17 (0.95 - 1.44)	0.88 (0.65 - 1.20)	0.91 (0.51 - 1.62)	1.00 (0.44 - 2.29)
Current Chronic Hypertension	65	3.74 (1.90 - 5.83) ***	2.03 (0.87 - 4.73)	1.19 (0.16 - 8.69)	2.64 (0.36 - 19.55)
Pregnancy Induced Hypertension	884	1.65 (1.34 - 2.05) ***	1.19 (0.87 - 1.63)	0.95 (0.50 - 1.81)	1.16 (0.48 - 2.80)
Anemia	213	1.43 (0.68 - 1.67)	0.87 (0.44 - 1.71)	1.09 (0.34 - 3.48)	2.50 (0.76 - 8.20)
Other Current Conditions	161	2.58 (2.02 - 4.46) ***	1.92 (1.10 - 3.37) *	6.33 (3.28 - 12.22) ***	n/a
Placental Anomalies	350	1.10 (0.87 - 1.43)	0.57 (0.30 - 1.07)	0.67 (0.29 - 2.05)	1.47 (0.45 - 4.83)
Fetal Clinical Wasting	163	4.09 (3.17 - 4.68) ***	2.71 (1.66 - 4.45) ***	19.80 (11.96 - 32.80) ***	1.03 (0.14 - 7.53)
Infant of Diabetic Mother	232	1.86 (1.30 - 2.66) ***	1.70 (1.04 - 2.80) *	1.35 (0.49 - 3.73)	0.71 (0.10 - 5.21)

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

significance determined from Chi-squared statistic

5.4 Results of Multivariate Analysis

The results of the bivariate analysis indicate that several variables have a significant relationship with at least one of the outcomes under study. The next logical step in the analysis was to combine these significant variables into a single model. Typically, each outcome under study in this thesis would be compared individually to normal births in a multiple logistic regression as in Burra, 2002 or Yang *et al.*, 2002 for example. As discussed in chapter three, this problem is multinomial in nature, with five broad outcomes. Using a MNL model, this can be modeled as 4 simultaneous equations comparing each of the outcomes to normal births. Further, it is also the case, that the model tests all possible combinations among the five groups although it only displays coefficients for the four comparisons. Table 5.5 displays the conditional odds ratios (CORs) that result from the MNL model that had a significance level of $p < 0.10$. Using a likelihood ratio test, all variables, except for being under the age of 19 and being admitted to hospital for a complication, made significant contributions to the overall model at the $p < 0.10$ level. The two variables that did not contribute significantly to the overall equation still produced significant coefficients for at least one of the outcomes and were retained in the model (refer to Table A6.2 in Appendix VI for the likelihood ratio tests).

Several variables yielded a significant relationship with preterm births. Being married resulted in a decrease in the relative odds of having a pregnancy result in a preterm birth. Smoking and prescription drug use resulted in increases in the relative odds

Table 5.5. Statistically significant CORs (p < 0.10) from the MNL model (n = 15800)

Variable	Preterm Births		Anomalies		LBW		Stillbirth	
Marital Status	0.87(0.76 - 0.99)	*	--		--		1.64(0.95 - 2.82)	
Mother Less Than 19 Years of Age	--		--		--		1.84(0.93 - 3.62)	
Smoking During Pregnancy	1.30(1.15 - 1.47)	***	--		2.40(1.71 - 3.37)	***	1.78(1.13 - 2.85)	*
Prescription Drug Use	2.70(1.88 - 3.88)	***	1.83(1.06 - 3.15)	*	--		3.81(1.24 - 11.69)	*
Nulliparity or Grand multiparity	1.52(1.34 - 1.72)	***	--		1.86(1.29 - 2.67)	**	1.98(1.17 - 3.33)	*
Previous Stillbirth	1.80(1.21 - 2.66)	**	--		--		3.68(1.36 - 9.94)	*
Previous Low Birthweight	2.04(1.65 - 2.53)	***	--		2.00(1.13 - 3.51)	*	1.86(0.87 - 4.00)	
Previous Adverse Outcome	2.27(1.83 - 2.82)	***	1.67(1.23 - 2.26)	**	--		2.82(1.28 - 6.23)	*
Admitted to Hospital with Complications	1.17(1.02 - 1.34)	*	--		--		--	
Chronic Hypertension	2.06(1.38 - 3.08)	***	1.38(0.76 - 2.52)		--		3.02(0.85 - 10.78)	
Pregnancy Induced Hypertension	1.52(1.28 - 1.79)	***	1.22(0.97 - 1.55)		--		1.76(0.94 - 3.28)	
Anemia	1.42(1.12 - 1.79)	**	--		--		4.39(2.42 - 7.98)	***
Other Complication (not including above)	2.15(1.64 - 2.82)	***	1.82(1.25 - 2.64)	**	6.92(4.37 - 10.96)	***	--	
Fetal Clinical Wasting	4.42(3.31 - 5.91)	***	4.96(3.46 - 7.10)	***	19.56(12.74 - 30.03)	***	--	
Infant of Diabetic Mother	2.40(1.94 - 2.96)	***	1.72(1.27 - 2.33)	***	--		--	
Living Within 3 Km of the Coke Ovens	1.42(1.11 - 1.81)	**	--		--		--	
Place of Residence			--		--		--	
- Sydney	--		1.38(1.03 - 1.85)	*	--		--	
- Industrial Cape Breton	--		--		--		--	
- Rest of Cape Breton County	--		--		--		--	
Log-Likelihood (Constants Only)	5532.79							
Log-Likelihood (Full Model)	4607.91							
rho-squared	0.167							

* p < 0.05 ** p < 0.01 *** p < 0.001

of having a preterm birth. Maternal age in any form proved to be insignificant, along with the deprivation score (omitted from tables since it is not statistically significant at $p < 0.10$). Controlling for pregnancy history was important as evidenced by the strength of significance of nulliparity or grand multiparity, prior LBW, other negative outcome ($p < 0.001$) and prior stillbirth ($p < 0.01$). Having a prior abortion resulted in an increased COR as well, however, it was statistically insignificant. All five variables used to represent current maternal diagnoses yielded statistically significant increases in the relative odds of having a preterm birth. Fetal clinical wasting produced the strongest relationship with preterm births. When a neonate presented with recent fetal starvation, the relative odds of having a preterm birth was approximately 4.5 times greater ($p < 0.001$), than when neonate did not have this condition. The other neonatal measure, being an infant of a diabetic mother, also increased the odds of preterm birth (COR: 2.40). Finally, the odds of having a preterm birth increased if the mother resided within three kilometers of the Coke Ovens site (COR: 1.42, $p < 0.01$). All other proximity measures failed to yield a significant result.

The number of significant variables for congenital anomalies is less than those for preterm births. In this case, prescription drug use, previous adverse outcome, pregnancy induced hypertension, other pregnancy complications (see Appendix V for definition), fetal clinical wasting, and infant of a diabetic mother all yielded a statistically significant increase in the odds of congenital anomaly. For the proximity measures, residing in Sydney versus the other parts of ICB and CBRM was also statistically significant (COR: 1.38, $p < 0.05$).

For LBW, even fewer variables were statistically significant in the MNL. Smoking, either being nulliparous or grand multiparity, having a prior pregnancy result in a LBW, other current pregnancy complications such as eclampsia and hyperemesis, and fetal clinical wasting. The latter variable produced the highest increase in the conditional odds in the entire analysis, where the COR was almost 20 times higher if the neonate had recent fetal starvation.

Significant CORs for stillbirths include smoking and prescription drug use, nulliparity or grand multiparity, having a prior stillbirth, having a prior LBW and having a prior adverse outcome (other than the aforementioned ones) and the mother having anemia. The other variables in the table were not significant at the $p < 0.05$ level in the MNL model. Marital status is the only variable that takes the wrong sign in the analysis however it is not statistically significant. All variables generated in the GIS were insignificant both for LBW and for stillbirths.

As mentioned previously, all variables but two contributed significantly to the model (See Appendix VI) with the overall goodness-of-fit as determined by ρ^2 equal to 0.167. The ρ^2 statistic is a measure of the information explained by the model (Ben Akiva and Lerman, 1985). Wrigley (1985) suggests that values of ρ^2 between 0.2 – 0.4 are a very good fit. The model, however, did a poor job in classifying the outcomes correctly (see Table A6.3 in Appendix VI for the Classification table).

Table 5.6 presents the results for the primiparous woman. There were considerably fewer significant variables in this case. For preterm births, smoking,

Table 5.6. Statistically significant CORs ($p < 0.10$) from the MNL model for primiparous women (n = 6009)

Variable	Preterm Births	Anomalies	LBW	Stillbirth
Mother Older Than 34 Years of Age	--	--	2.63(1.08 - 6.41) *	--
Smoking During Pregnancy	1.30(1.09 - 1.56) **	--	1.64(1.03 - 2.63) *	--
Prescription Drug Use	3.19(1.81 - 5.60) ***	--	--	--
Deprivation Score				
1st Quintile (Most Deprived)	--	--	--	--
2nd Quintile	1.27(0.97 - 1.66)	--	--	--
3rd Quintile	1.35(1.01 - 1.80) *	--	--	--
4th Quintile	--	--	--	--
5th Quintile (Least Deprived)	--	--	--	--
Chronic Hypertension	2.11(1.11 - 4.00) *	--	--	--
Pregnancy Induced Hypertension	1.49(1.19 - 1.87) ***	--	--	--
Fetal Clinical Wasting	4.00(1.19 - 1.87) ***	4.32(2.56 - 7.30) ***	29.77(17.29 - 51.26) ***	--
Infant of Diabetic Mother	1.84(1.27 - 2.66) ***	1.86(1.12 - 3.09) *	--	--
Log-Likelihood (Constants Only)	1123.34			
Log-Likelihood (Full Model)	870.16			
rho-squared	0.225			

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

prescription drug use, chronic hypertension, pregnancy-induced hypertension, fetal clinical wasting and infant of a diabetic mother all increased the relative odds of having a preterm birth. Additionally, residing in an enumeration area that was in the mid-range of deprivation produced a statistically significant increase as well. For congenital anomalies, only two variables were significant. Those were fetal clinical wasting and being an infant of a diabetic mother. For LBW, maternal age being greater than 34 years and smoking were significant, as was fetal clinical wasting which produced a COR of approximately 30. None of the variables were statistically significant for stillbirths. All variables contributed significantly to the overall model except for maternal age of 35 and over, chronic hypertension and the deprivation index (see Table A6.4 in Appendix VI). The overall model performance was better than for the full model with $\rho^2 = 0.225$. The model however did a poorer job of properly classifying the outcomes than the full model (see Table A6.5 in AppendixVI).

First-order interaction terms were constructed for all statistically significant variables and entered into each model in a stepwise fashion. None of the variables constructed contributed to the model in a meaningful way.

5.5 Discussion

In general, the results of sections 5.3 and 5.4 suggest that proximity to the Tar Ponds/Coke Ovens site did not have a major impact on whether a woman experienced an adverse reproductive outcome. An exception to this was found in two instances where statistically significant associations were found with location. The conditional odds of having a preterm birth increased significantly if the location of residence was within 3

kilometers of the Coke Ovens site and the conditional odds of having a congenital anomaly were significantly higher in Sydney versus the other parts of CBRM and ICB. These findings were produced while controlling for the suite of predictors presented in Table 5.5. Straight-line distance failed to make a significant contribution to any of the models as a proxy for exposure.

The result for preterm births for the full data set (n=15800) matched the finding produced by Burra (2002), who used a cross-sectional survey of 500 women to examine the relationship of proximity to the site and reproductive health. However, similar to Burra (2002), this result must be interpreted with caution. Although key variables such as smoking, several aspects of reproductive history and various pregnancy-related complications are controlled for, variables representing SES and education are missing. As Kramer (2003) points out, both of those factors are important determinants of preterm birth, since women of lower SES and education often have lower intakes of essential nutrients required for successful parturition. Kramer (2003) adds that low SES and education are associated with many adverse behavioural and psychosocial factors associated with adverse outcomes such as preterm births. The Atlee Perinatal Database, while very strong on the biomedical aspects of reproductive health, does not collect any information pertaining to SES. With that in mind, a deprivation index was created from the 1996 Census (see chapter 3) and included in the model in an attempt to capture the influence that parental SES and education has on reproductive health. This variable failed to produce a significant result no matter what form it took. Unlike Vrijheid *et al.* (2000), who found a statistically significant gradient in the risk of congenital anomalies

as one moved from the least deprived to the most deprived quintiles of the distribution, this study failed to demonstrate any gradient, regardless of significance. This result was likely produced by the lack of variability in the measure due to the grouping of enumeration areas into deprivation quintiles. Although it is standard practice to do so, in this case, with only approximately 170 enumeration areas, not enough variability was present to detect any significant differences. In addition, Vrijheid *et al.* (2000) utilized a case-control approach to their study, where they had 858 cases of congenital anomaly, and 1764 cases of non-malformed control births. In this study, the majority of the population is being used (minus the observations with missing data), and as such, there is not as much variation to detect a significant gradient since the majority of events are normal births.

The second association with location and an adverse outcome occurred with congenital anomalies and residence in Sydney versus ICB (excluding Sydney) versus the remainder of CBRM. This result agrees with the findings of Dodds and Seviour (2001), where they found a 25% increase in the rates of congenital anomalies in Sydney versus the rest of Nova Scotia while controlling for maternal age, smoking and parity. A gradient was observed in the contingency analysis whereby residing in Sydney always had the highest OR regardless of outcome, followed by residing in ICB (excluding Sydney) and finally residing in CBRM, however, this was not statistically significant and the gradient did not hold in the MNL model, as in some instances residing in ICB had a higher OR than residing in Sydney.

Although these two findings suggest that living closer to the Tar Ponds/Coke Ovens site may be important with regards to preterm births and congenital anomalies, it is possible that there is residual confounding associated with SES and maternal education. However, as Burra (2002) points out, it is also possible that exposure (as measured by proximity to the point source) to the contaminants in the physical environment may be exerting an influence. However, the lack of variables to capture the parents' educational attainment and SES remains a limitation to the study. Neither of the two findings was found to be significant in the model dealing with primiparous women.

With regards to the proxies for exposure, it is important to note that they are limited. In an ideal situation, the researcher would like to measure a life-long dose to an exposure, but this is never possible. Instead, researchers must resort to other measures to try to estimate individual level exposure. For the most part, the methodology chosen has been driven by readily available data or by previous practice. Methods include indirect monitoring techniques such as the proximity measures incorporated in this research, dispersion models, and interpolation based on a stationary monitoring network; and direct techniques like personal monitoring devices. Recently, studies have been undertaken to determine the performance of these measures. Briggs (2000) notes that interpolation performs poorly in comparison to dispersion models and that proximity measures are the poorest measure available of the three. Since there are not enough monitoring stations to produce a stable surface, interpolation had to be ruled out. As pointed out in chapter three, the primary point source of pollution ceased operation in 1988 rendering dispersion

modeling as ineffective, thus, leaving proximity to the contaminated areas of the Tar Ponds/Coke Ovens site as the only measure of exposure available for the analysis.

The limitations of the geocoding had a role to play in the proximity measures as well since several records were geocoded to identical locations. This meant that the variation in the distance values was decreased, since several observations had identical distances associated with them. This potentially led to the ineffectiveness of the straight-line distance as a proxy of exposure. The zonal based concentric-ring based measures were more reliable since all observations were correctly coded to the community of residence.

With respect to the other covariates, all performed in the expected way with few exceptions. With regard to preterm births, the OR for marital status took the expected sign, however, the magnitude of the effect was less in this study than other studies such as Goldberg *et al.* (1995) and Kramer *et al.* (1998). The OR for stillbirth was elevated for married women, but not statistically significant at the 0.05 level. This result matches the findings of Forssas et al. (1999). Maternal age, recognized as being an important predictor of all adverse reproductive outcomes (Kramer *et al.,* 2002, Moutquin, 2003) had little impact in this analysis, showing only a marginal effect for stillbirths. However, when examining primiparous cases, maternal age greater than 34 was important for LBW, more than doubling the odds of the occurrence of that event. Smoking was significant for all outcomes aside from congenital anomalies, taking the appropriate sign in each case and showing some consistency in magnitude of effect with the literature. For example, Stephansson *et al.* (2001) found that smoking increased the odds of

stillbirth by 70%, whereas this study observed a 78% increase in odds. Vassilev *et al.* (2001) and Tough *et al.* (2001) reported an almost doubling of the odds of LBW and preterm births respectively when the mother smoked, while the ORs in this study were 2.4 (95% CI. 1.71-3.37) for LBW and 1.30 (95% CI. 1.15-1.47) for preterm births.

Covariates relating to medical risks predating the current pregnancy performed as expected. Nulliparity and grand multiparity were associated with elevated ORs for preterm birth, LBW and stillbirths, almost doubling the odds of the last two outcomes. Kramer *et al.* (1998) found nulliparity to be a predictor of preterm birth, although the magnitude was less than what is reported in this study. Gouveia *et al.* (2004) found that nulliparous women had increased odds of LBW, while Butler and Kalasinski (1989) report a significant increase in the OR for stillbirths for grand multiparity (OR: 1.68, 95%CI: 1.19-2.17). In this study, a prior stillbirth increased the odds of having another stillbirth by 3.5 times. There was corroborating evidence for this from Forssas *et al.* (1999) and Surkan *et al.* (2004) who both found an effect of the similar magnitude in other populations. The fact that a prior stillbirth was associated with a preterm birth in this population is corroborated by Robson *et al.* (2001) who found a similar association in a retrospective study of women drawn from a South Australian birth registry. Previous LBW was found to be an indictor of subsequent preterm births, LBW and stillbirths in this population. All of these links have been established in other populations as well (Surkan *et al.*, 2004).

Present pregnancy maternal diagnoses proved to be important for this study also. For preterm births Moutquin (2003) cites that pregnancy-induced hypertension is related

to preterm births since women with severe cases are often induced into delivering early. Although the number of pregnancies induced prematurely in this population is unknown, the models both indicate that pregnancy-induced and chronic hypertension are important covariates of preterm birth. The ORs for congenital anomalies and stillbirths are also elevated for these covariates, however, the result is not statistically significant at the $p < 0.05$ level as can be seen in the confidence intervals. The relationship observed between stillbirth and anemia is interesting since it is counter to the recent evidence in the literature. Stephansson et al., (2000) report that while anemia is related to other outcomes such as preterm births (a relationship demonstrated in this study also), no such relationship was found with stillbirth. Using a large population-based registry in Sweden, Stephansson *et al.* (2000) found that high concentrations of hemoglobin were associated with stillbirth rather than anemia. In this study, anemia was significant in the contingency analysis as well as in combination with several other predictors in the MNL. Further, the magnitude of the effect was consistent in the bivariate and multivariate analyses, thus, ruling out multi-collinearity. This suggests that anemia may be important in this population with regards to stillbirth, however, it is important to note that there were a small number of pregnancies that resulted in stillbirth. The relationship between stillbirth and anemia may be worthy of further study, perhaps using a case-control approach, where twice as many non-stillbirths would act as a representative set of controls.

Two neonatal measures were important in this study; fetal clinical wasting and being an infant of a diabetic mother. Fetal clinical wasting (also known as recent fetal

starvation or fetal malnutrition) was positively associated with preterm birth, congenital anomaly and LBW. The magnitude of the OR for LBW was larger than any other effect in the study. This effect however, was likely an artifact of the coding of preterm births and LBW. Kramer (2003) identifies that birth weights are determined by two processes, gestational age and rate of fetal growth. Gestational age is a non-factor for LBW in this study, since all observations with less than 37 weeks gestation were already coded as preterm. This means that the LBW variable as coded in this study is largely impacted by the rate of fetal growth. This rate of fetal growth can be affected by malnutrition, thus, becoming growth restricted *in utero*. The definition of fetal clinical wasting used in the Atlee Perinatal database is based on the work of Scott and Usher (1966), who describe it as a combination of being small-for-gestational age and those who show clinical signs of soft tissue wasting. This can be the result of placental insufficiency or dysfunction and postmaturity (Scott and Usher, 1966). Since the variables were coded to ensure mutually exclusive categories, it is possible that the value of the OR for fetal clinical wasting with regards to LBW is not different from other populations; however, future research would be required to test that hypothesis. The other neonatal measure, being an infant of a diabetic mother was found to elevate the odds of preterm birth and congenital anomalies. Diabetes is recognized as being an important predictor of stillbirth (Forssas *et al.,* 1999), however, in this population it did not contribute meaningfully to the model.

A limitation to studies of this nature, discussed by Burra (2002), was the lack of independence in the observations of models of this type. In that research, the results from a standard logistic regression model were compared to a logistic-normal mixed model

with a random intercept. This type of modeling was considered for this data set, but was not applied for two reasons. The first issue is that not every pregnancy by each woman in the database is included in this data set. There are instances where the parity variable for a woman indicated that her final entry into the database corresponded to her eleventh pregnancy, however as can be seen in Table 5.7, the maximum amount of times any woman appears in the database is six. There are two reasons for that. In some instances, a woman's final entry into the Atlee was near 1988 when the database started. For example, that could have been a woman's fifth pregnancy, but the prior ones would not be recorded in the Atlee. What would have been recorded for that woman however, was whether or not she gave birth in the past to a LBW baby, had a stillbirth, or had an abortion (spontaneous or induced). The second reason for missing pregnancies was due to missing data (described in chapter three). One woman for example had information for her first, second, and fifth pregnancies. The information for the other two pregnancies could have been lost due to missing maternal identifiers, parity information, some other key variable or if one or more of the pregnancies resulted in a multiple birth.

Table 5.7 Number of pregnancies contributed to the Atlee database by each woman (1988 – 2002)

Number of Times Appeared in Atlee	Number of Women	Total Number of Pregnancies
1	6836 (63%)	6836 (43%)
2	3146 (29%)	6292 (40%)
3	728 (7%)	2184 (14%)
4	96 (.89%)	384 (2.4%)
5	16 (.15%)	80 (.51%)
6	4 (.04%)	24 (.15%)
Total	**10826 (100%)**	**15800 (100%)**
Average number of pregnancies contributed by each woman:		15800/10826 = 1.46

The second issue that precluded the use of a random-effect model is related to the way in which these random-effects are implemented. Typically, data of this type need to be modeled as panel data (Paez, 2004, personal communication). In a panel data setting, usually there is a time-series of observations on multiple entities, such as individuals. The modeling of reproductive outcomes is an analogous situation, where instead of repeated measures on a woman for specific time intervals, the repeated measures are for pregnancies for each woman. The problem with implementing panel data in this setting is tied to the missing records. In a panel setting each woman should have the same number of pregnancies occurring. As can be seen in Table 5.7 that is not the case, with the majority of women contributing one or two pregnancies, but the maximum number contributed is six. Setting these up in a panel would result in missing data for several of the pregnancies and the estimation would be biased substantially (Arellano, 1995). Further, as highlighted previously, there is incomplete data per woman, since some of the pregnancies contributed may have been lost to coding errors, or missing information. The analogy to time-series is that some of the time intervals in the middle of the study period would be missing. Although there are methods to deal with a lagged dependent variable (i.e. unbalanced panel data, see Arellano and Bond, 1991), methods do not exist to adequately accommodate missing time intervals in the middle of a study period.

Burra (2002) found little difference between the standard models and the ones incorporating random effects. This result was similar to Butler and Kalasinski (1989). Both Butler and Kalasinski (1989) and Watier *et al.,* (1997) conclude that the negligible

differences were due, at least in part, to the small number of pregnancies per woman. In Burra (2002), the average number of pregnancies was 3.3 per woman, while Butler and Kalasinski (1989) reported the average number of births per woman in their study to be 1.3. As can be seen in Table 5.7, the average number of births in this study is 1.46. Although this is in between the averages of Burra (2002) and Butler and Kalasinski (1989), and both found a similar negligible impact in their respective analyses, it cannot be concluded with certainty that the same result would occur here. However, given the limitations with the data, the only option was to apply the standard MNL model controlling for several aspects of reproductive history. This was also the reason for modeling the data with only maternal first pregnancy.

As previously noted, the covariates that were statistically significant all performed in the expected way according to the literature. The models however, were poor in properly classifying the outcomes (see tables A6.3 and A6.5 in Appendix VI). The classification would have likely been improved in that regard with the introduction of additional covariates such as SES and psychosocial stress. With regard to preterm birth, for example, Kramer *et al.,* 2001 point out that most cases occur without any known cause. But they do mention the importance of SES and psychosocial stress. They hypothesized that chronic and acute psychosocial stressors can lead to increased secretion of placental corticotropin releasing hormone which has been found to stimulate premature contractions of the uterus (Kramer *et al.,* 2001). They also note that the erosion of personal resilience and the lack of empowerment, that is rampant in the socially disadvantaged (as highlighted in chapter 2), are important for preterm birth as

well. In section 2.2 of chapter 2, the determinants of population health were listed. Of the twelve components listed there, the Atlee only collects information on two of them, personal health practices and biomedical variables. The remainder, which include income and social status; social support networks; education and literacy; employment and working conditions; social environments; physical environments; health services; gender; and culture are not represented in the Atlee Perinatal database, thereby representing a limitation to the study. In a companion thesis, Burra (2002) was able to include variables from these other components that make up the determinants of health since she used a primary data source that consisted of a survey for 500 women in the Sydney area. This study was limited in that a secondary data source was utilized. However, it is important to note that while Burra (2002) was able to find a significant relationship with preterm birth and proximity to the Tar Ponds, this thesis found the same relationship using a different modeling framework and variables that were not accounted for in her thesis, thus making these two bodies of research complementary.

Another explanation for the classification issue lies in the way the outcomes were coded at the outset. The classification for congenital anomalies could have been improved by breaking down the variable into more specific anomaly types as in Vrijheid *et al.,* (2002b). While this would have improved the specificity of the outcomes, this would have only served to split the number of observations in the dependent variable into smaller groups, thereby decreasing the statistical power of the model. Dodds and Seviour (2001), while examining specific anomalies in their research, also examined all major anomalies combined. The dependent variable in this thesis was constructed to match

their work so that comparisons could be made between the two bodies of research. The poor classification of stillbirths and LBW was due in large part to the small numbers of those two outcomes in relation to the other categories in the dependent variable.

It is important to note that spatial autocorrelation was not accounted for in this research. The primary distinguishing property of spatial data is that neighboring data samples tend to systematically affect each other or some underlying variable or cause is missing from the model estimation. Thus, the classical assumption that data samples are generated from independent and identical distributions is not valid. It is possible that the observations are spatially dependent either in the outcomes themselves, or in the error terms. The reason that autocorrelation was not accounted for here was due to limitations with the data. Mohammadian *et al.,* (2001) describe a spatial multinomial logit model, where a parameter designed to capture the autocorrelation is included in the utility term. To account for spatial autocorrelation, a matrix would need to be generated to capture the spatial locations of pregnant women. Given the limitations in the geocoding (see Chapter 4), several women were geocoded to the identical location, meaning that the resulting covariance matrix would not be able to invert making model estimation impossible. The idea of distributing the identically located women randomly was also explored as a way around this issue. However, it is possible that autocorrelation could be induced in the data set artificially through methods such as this, thus biasing the estimation. In other research, autocorrelation has been specified amongst the alternatives as in Ben Akiva and Bolduc, (1996) for example. In this research, there is little theoretical basis for implementing an autocorrelation parameter in this way. For example, a woman with an

adverse outcome will not exert an influence on a neighbouring woman in such a way that the same adverse outcome will be induced in her neighbour. What is more likely is that the error terms are autocorrelated, signifying that there is a missing variable(s) that varies spatially. Aside from the model described by Mohammadian *et al.,* (2001), MacMillen (1992) describes a probit model, which specifies a parameter designed to capture spatial autocorrelation in the error terms. This model, while only configured for binomial outcomes, may potentially be extended to handle a multinomial outcome and is something to consider for future research. However, in order for either model to function properly, the data limitations would need to be overcome.

5.6 Summary

In presenting the results of the bivariate and multivariate analyses, this chapter has addressed the second objective of this research: to determine if the pattern of adverse outcomes is related to proximity to the Tar Ponds/Coke Ovens site while controlling for several risk factors. For preterm births, residing within three kilometers of the site led to a statistically significant increase in the COR, while controlling for several biomedical covariates. Also, residing in Sydney had a significantly elevated COR for congenital anomalies, while the same effects could not be seen in ICB (excluding Sydney) and the rest of CBRM. However, several key covariates such as SES and psychosocial factors were missing due to the limitations of the secondary data source employed in this research. The implications of this as well as the other findings of this thesis, the contributions of this research and the directions for future research are presented in the following chapter.

CHAPTER SIX

CONCLUSION

6.1 Introduction

This thesis presents the findings of a spatial study designed to explore the link between contamination in the vicinity of the Tar Ponds and reproductive health of women in Sydney, Nova Scotia and the surrounding area. This research is one component of a larger programme designed to explore the impacts of contamination in the Muggah Creek Watershed on reproductive and psychosocial health of Sydney residents. Two objectives have been addressed by this research:

i. To assess the spatial pattern of various types of adverse reproductive events, plausibly linked to the environmental exposure of interest;
ii. To determine if this pattern is related to proximity to the Tar Ponds/Coke Ovens site.

This chapter summarizes the key findings with respect to each of these objectives. Following the summary of findings, the contributions of this research are discussed. The chapter concludes with recommendations for future research.

6.2 Summary of Findings

Point pattern analysis was used to address the first research objective. Observations from the Atlee Perinatal Database were geocoded at the Nova Scotia RCP using a variety of methods (see Table 3.1). These geocodes were improved using address matching where possible, with the remainder geocoded with a PCCF from DMTI Spatial. Original coordinates were retained for any records that did not geocode by either of these

methods (see Table 4.1 for results). The results were then mapped (section 4.2.2) and analysis was conducted to assess the spatial patterns of the outcomes under study.

In order to assess clustering on a global-scale, the bivariate K-function was applied to the data (4.3.1). The point patterns used in the visualization phase of the analysis represented the cases, while controls were selected from all normal live births. This statistic was computed for the entire 15-year period (1988-2002) as well as for three five-year intervals (1988-1992, 1993-1997, 1998-2002). The majority of the results indicated weak clustering (i.e., positive values of the test statistic, but statistically insignificant as determined through Monte Carlo simulation) with three cases indicating weak repulsion (preterm births and LBW for the 1998-2002 temporal subset and anomalies for the 1988-1992 subset). Two cases showed significant clustering at some spatial scales (preterm births for the first two temporal subsets). An analysis for stillbirths was conducted only for the full time period due to the small number of cases and failed to produce a significant result (Table 4.2).

The results described in section 4.3.2 examined clustering on a local scale. To determine where clusters might be located, the variation in the mean of the process was examined using a ratio of kernel estimates with a 1 km bandwidth applied to each of the outcomes (in both the numerator and the denominator). For preterm births, the mean of the process varies across space, with the majority of estimates between 50 and 150 preterm births per 1000 live births. The pockets of estimates greater than 200 preterm births are all areas where there was more than one adverse event geocoded to an identical location, divided by very few normal events in the denominator. Significance of clusters

was tested using 99 simulations, computed under the random labeling hypothesis (see Chapter 3). Each of the grid cells in figure 4.11 satisfied the condition that the observed values were greater than two standard deviations away from the mean value of the 99 simulations. The areas closest to the Tar Ponds displayed levels of 50 to 150 preterm births per 1000. This pattern, while different than the results for Cape Breton County, was similar to the other areas of ICB, both in coverage (i.e., most of the respective administrative areas are covered by elevated estimates) and in magnitude (mostly 50 to 150 events per live births).

The pattern for congenital anomalies was similar to that of preterm births, where much of ICB displayed elevated levels of anomalies per 1000 births. In figure 4.13, there appeared to be a grouping of cells in the 100-150 classification north of the Tar Ponds in Whitney Pier. These elevated levels were consistent throughout much of the neighbourhood. To the south of the Tar Ponds in Ashby, the estimates were consistently lower (ranging from 50-100 anomalies per 1000 births). Further south, in Hardwood Hill and South End (southern part of Sydney) the levels were elevated again (100-150 anomalies per 1000). In the other communities of ICB, the predominant level of anomalies per 1000 births was between 50 and 150.

The results for LBW and stillbirths were quite different from the previous two outcomes both in coverage and magnitude. This was due to the smaller numbers of events in the numerator (n=185 and n=100 respectively). There are very few estimates formed in Sydney Mines and North Sydney (to the northwest), however, significant clusters were exhibited in New Waterford and Glace Bay (Figure 4.15). The magnitude

of the estimates have decreased in comparison to the previous two outcomes with the mode equal to 10-20 events per 1000 live births. Within Sydney, much of Whitney Pier did not have an estimate of LBW per 1000 live births, while North End, Hardwood Hill and parts of Ashby did exhibit some elevated levels. Stillbirths had the lowest estimates per 1000 births. The mode of this outcome was the same as for LBW (10 – 20 per 1000 births). Within Sydney, there is a pocket of estimates in Whitney Pier, but the majority of estimates occur south of the Tar Ponds in Hardwood Hill (Figure 4.16). In the other communities of ICB, the coverage is not as great as several areas did not have any estimates.

This component of the research was useful in that it helped to confirm the findings of Dodds and Seviour (2001), where they found a slight increase in congenital anomalies in Sydney versus the region. It also extended their work by pinpointing areas within not only Sydney, but the remainder of ICB that were elevated over the background rate, not only for congenital anomalies, but for all of the outcomes under study. In addressing the first objective, the results of the bivariate K-function demonstrated that for preterm births, there was clustering at some spatial lags on a global scale. The kernels indicated that the spatial pattern of adverse reproductive outcomes was not random, especially for preterm births and congenital anomalies since all of Sydney and the majority of ICB had elevated estimates of these outcomes per 1000 births. The results for LBW and stillbirth were also non-random however the pattern was not as dominant as it was for preterm births and congenital anomalies.

To address the second objective, contingency analysis was conducted with each outcome under study and various measures of proximity to the Coke Ovens site. A GIS was used to construct various proxies for exposure, including: concentric rings of a one-kilometer radius each around the point source as in Bhopal et al (1999); geographic groupings by community (Sydney, rest of ICB, rest of CBRM) as in Guernsey *et al.* (2000); straight-line distance from point source to observations as in Dolk *et al.* (2000). For the distance-based zones (i.e., concentric zones around the Coke Ovens) there was not a clear gradient of decreasing odds ratios with increasing distance away from the Coke Ovens site for any of the outcomes (Table 5.1). The only statistically significant relationship was with preterm births and residing in the zone 2-3 kilometers away from the Coke Ovens site (OR = 1.23 (95% CI: 1.02 – 1.48)). The sum of the zones was also examined (i.e., from 0 – 3 kilometers) yielding a statistically significant relationship with preterm births (OR = 1.19 (95% CI: 1.05 – 1.35)). The results of the contingency analysis for place of residence and an adverse outcome produced a gradient in the ORs with Sydney having the highest values, followed by ICB, and finally the rest of CBRM. It should be noted however that the only statistically significant result from this analysis was for preterm births, where the odds of having an adverse event are significantly lower in the more rural areas of the county. Straight-line distance was examined using a t-test to compare the mean distance of the cases versus the controls. Table 5.2 demonstrates that there was a statistically significant difference between the two groups for stillbirths, with the cases having a shorter mean distance to the coke ovens site than the non-cases, thus, lending support to the spatial portion of the study where a pocket of elevated

estimates was found in Whitney Pier. All other results were insignificant. The entire analysis was repeated using first pregnancies only, however, no statistically significant results were observed.

Contingency analysis was performed on several variables obtained from the Atlee Perinatal Database or from the 1996 census (Table 5.3). Several significant associations were observed using a chi-squared test for covariates associated with maternal characteristics, pregnancy history, present pregnancy maternal diagnoses, and neonatal measures. The same analyses were carried for the primiparous women (n = 6009) in this data set and produced a fewer number of significant associations (Table 5.4).

All significant covariates from the bivariate analysis were combined in a multivariate model with a multinomial dependent variable for both the full data set and the primiparity data set. The respective goodness-of-fit measures for each of the models were 0.167 and 0.225. Both models, however, performed poorly in terms of correctly classifying the outcomes (see Appendix VI). However, several significant associations were exhibited in the full MNL model with each of the outcomes under study (Table 5.5). The exposure variables yielded a significant result in two instances. For preterm births, residing within three kilometers of the Coke Ovens site yielded a significant result while controlling for several covariates. All covariates performed as expected in the model, taking the correct sign. The only other proxy that yielded a significant result was residing in Sydney versus the other communities in ICB and CBRM and congenital anomalies. Straight-line distance did not yield a significant result for any of the outcomes. The

results for primiparous woman exhibited fewer significant relationships (Table 5.6), with the exposure proxies not contributing significantly to the model.

The result for preterm births for the full data set (n=15800) matched the finding produced by Burra (2002), who found proximity to contribute significantly to the model. However, similar to Burra (2002), this result must be interpreted with caution due to several limitations with the data set. Although key variables such as smoking, several aspects of reproductive history and various pregnancy-related complications are controlled for, variables representing SES and education are missing. Efforts to capture SES and maternal education through a deprivation index proved fruitless.

The result for congenital anomalies and place of residence (Sydney versus ICB versus CBRM) agreed with the findings of Dodds and Seviour (2001), where they found a 25% increase in the rates of congenital anomalies in Sydney versus the rest of Nova Scotia while controlling for maternal age, smoking and parity.

These results suggest that the contamination of the Muggah Creek Watershed may have played a role in affecting reproductive health adversely. However, these results are inconclusive due to the limitations with the data set, exposure variable, potential lack of independence in the data, and the fact that spatial autocorrelation was not accounted for. As a result, a causal association cannot be established on the basis of this analysis. Further, a biological mechanism cannot be established either, since very few components that make up the determinants of health were captured in this secondary data set.

6.3 Research Contributions

This research makes several contributions to the literature. The methodological contributions of this thesis relate to the use of GIS and spatial analytical methods, and the modeling approach adopted in this thesis. GIS was used in all phases of this analysis, whether it be for plotting the distribution of adverse events, examining potential clusters of outcomes, or for the creation of proxies for exposure to be used in multivariate modeling. The use of GIS in health is now well established (Gatrell and Senior, 1999), however, very few reproductive health studies have adopted the use of GIS (Dolk, 1999). As such, this thesis (and the larger research programme) provides an example of the utility of GIS in addressing research questions involving reproductive health.

The use of spatial analytic methods, (chapter four) to determine if adverse outcomes tend to cluster in the Muggah Creek watershed is unique. A study of this nature has never been conducted in Sydney (nor in Canada) and the results of this phase of the analysis add to the body of literature surrounding the health effects of the area. In addition, the use of these methods contributes to the literature surrounding small area statistics, spatial epidemiology and reproductive health, since very few studies have been able to apply spatial methods at an individual level to examine adverse outcomes of pregnancy. For example, aside from Dolk *et al.,* (1998b), the bivariate K-Function has not been applied in reproductive health research. This thesis provides a second example as to the usefulness of that tool to address questions dealing with whether clustering existed on a global scale. With regards to localized clustering, methodological improvements were made to the spatial analytic tool proposed by Reader (2001). Using a

ratio of kernel estimates as opposed to the ordinary kernel employed by Reader (2001) allowed for the correction of the underlying population, which as Gatrell (2002) points out is critical since clustering will occur *a priori* in populated areas. The result of this correction is an estimate of the number of events per unit population as opposed to the number of events per area.

The use of GIS to create proxies of exposure represents another methodological contribution of this research. While Beyea and Hatch (1999) and Briggs (2000) demonstrate that GIS has been applied in exposure estimation for some time, it has rarely been used in studies involving reproductive health (see Vassilev *et al.*, 2001, Elliott *et al.*, 2001 and Burra, 2002). For this research categorical groupings and continuous measures of distance were used as proxies for exposure to assess the intra- and inter-urban variations in the health outcomes of interest.

A final methodological contribution lies in the modeling approach adopted in this research. The multivariate modeling employed in this analysis is categorical in nature, specifically making use of MNL models. While models of this type have been employed in studies dealing with discrete choices for the type of reproductive health care sought after (Glei and Goldman, 2000), an extensive search of the literature failed to produce any past research that has applied these methods in studies dealing with adverse reproductive health outcomes and the environment, so this work adds to the current body of literature (both for reproductive studies, and for environment and health studies where the range of possible outcomes may be multinomial in nature).

Substantively, this analysis helped to elucidate the spatial pattern of adverse reproductive outcomes around a point source of pollution, and research of this type formed a component of JAG's site assessment process. This thesis described the first attempt at employing spatial analysis to assess patterns of adverse reproductive outcomes in Sydney. In addition, it is one of the only studies that used spatial analysis as a tool to answer questions about reproductive health, rather than using reproductive health data to demonstrate the utility of a newly devised method as in Reader (2001) or Rushton and Lolonis (1995).

Another contribution made to the literature lies in the unit of analysis. Similar to Burra (2002), this research relies on the individual rather an aggregate measure such as a standardized ratio. Aside from Burra (2002), the only other reproductive health study in Sydney (Dodds and Seviour (2001)) relied on aggregating outcomes to establish rates of anomalies. As Walter (1991) demonstrates, there are several deficiencies when modeling with ecologic-level data including ecologic fallacy and the modifiable areal unit problem. Deficiencies such as these are not an issue when modeling with individual-level data, hence the broader movement towards analysis of this type in health geography. Another important contribution relating to the work of Dodds and Seviour (2001) is that this research corroborates their findings with respect to congenital anomalies using a different unit of analysis and a different approach to the modeling.

Finally, this research forms one of the substantive contributions to part (iv) of the larger research project. In part (iv), data from the other components were integrated into a GIS and spatial analysis was conducted. While useful in other aspects of the project, it

was this research that derived the greatest utility from a GIS, capitalizing on its functionality wherever appropriate to analyze the data in the Atlee Perinatal Database. The potential areas of future research are discussed below.

6.4 Future Research Directions

As outlined above, this research has made several contributions to the literature. Using a large secondary dataset such as the Atlee Perinatal Database in a cross-sectional setting proved useful for determining the prevalence of the various outcomes under study. If there is a criticism to make about a study of this nature, it is that performs relatively poorly when studying rare phenomena. In this study, stillbirths are comparatively rare events (n=80) and as such the results of the MNL model must be treated with caution. The association between stillbirth and anemia was an interesting finding considering that recent evidence suggests that it is high levels of hemoglobin rather than anemia that are associated with stillbirths. One avenue for future research is to re-examine this relationship in a case-control format. As explained by Mann (2003), having a large population-based database can be used to construct suitable controls to be combined with stillbirths in a case-control framework. In the same way, a case-control study can be applied for each of the outcomes to determine if the deprivation index performs in a similar manner as in Vrijheid *et al.* (2000), who observed an increase in risk of having a congenital anomaly with increasing deprivation.

As highlighted in Burra (2002) improvements to the exposure assessment is another area of further investigation. For example, the results of the soil samples

conducted recently in Whitney Pier would be useful to determine how much of the pollutants in the Tar Ponds are transported to the residents via deposition from the water cycle. Additionally, since several of the chemicals in the area are persistent, soil samples may also shed light on the amounts of pollutants that local residents may still be exposed to in their immediate surroundings.

One of the limitations in the MNL model is the inability to capture the impact of spatial dependence in the model. A few possibilities exist for future research in this area. One option is to use proportions per zone as the dependent variable rather than individuals. This would eliminate the problem of having several events occurring at one location, so the methods of Mohammadian et al. (2003) could be applied. Also, if the method proposed by McFadden (1992) could be extended to accommodate a multinomial dependent variable, which could be applied in this setting. If models were to be run in this fashion, then variables could be obtained from the census to control for SES and education. One caveat to this approach, however, is that through the aggregation involved to create the proportions, all limitations of modeling with ecologic-level data would be introduced. However, the proportional model (minus the spatial effects) could be compared to the individual model in this thesis to determine if there was general agreement between the two, and then the spatial effects could be included. Another possible solution would be to analyze only those records that were address geocoded (n=3008). In this way, the spatial effects can be estimated using the model specification described by Mohammadian *et al.* (2003) since all women will have unique locations.

Finally, with regards to the point pattern analysis conducted in chapter 4, the application of an adaptive kernel estimate could be an area of future research. This type of kernel is preferable since the number of events is held constant for each estimate and the bandwidth varies in order to capture the number of events specified by the user. In this way, the variance in the estimation process is consistent throughout the map. This was not applied in this case due to the limitations in the geocoding. If the geocoding was improved (see chapter 4) then adaptive estimation could be applied since several events will not be located at the same location (for example, there is more than 100 events at one location in Marion Bridge, 20 km south of Sydney). If the number of events at a unique location is greater than the amount specified by the user to be included in the estimate, then the bandwidth will be zero and all events will receive the maximum weight. Situations such as this introduce variation into the estimates, which is exactly what adaptive estimation is meant to avoid. A pilot study could be conducted on the 3008 observations that were address geocoded. In a study such as this, the research could compare the results from a fixed-width bandwidth versus an adaptive bandwidth.

BIBLIOGRAPHY

ACOG (2001). ACOG Practice Bulletin No. 31: Assessment of Risk Factors for Preterm Births. *Obstetrics and Gynaecology*, 98(4): 709-716

Agius, R (2001). What is Environmental Health? http://www.agius.com/hew/resource/envhlth.htm. Accessed on June 10, 2004.

Agresti, A (1996). An Introduction to Categorical Data Analysis. John Wiley & Sons INC., Toronto.

Arellano, M and Bond, SR (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies*, 58: 277-297.

Arellano, M (1995). On the testing of correlated effects with panel data. *Journal of Econometrics*, **59**:87-97

ATSDR, 2003. Toxicological Profile Information Sheet. http://www.atsdr.cdc.gov/toxpro2.html. Accessed on June 7, 2004

Baibergenova, A, Kudyakov, R, Zdeb, M, Carpenter, DO (2003). Low Birth Weight and Residential Proximity to PCB-Contaminated Waste Sites. *Environmental Health Perspectives*, 111(10):1352-1357.

Bailey, TC and Gatrell AC (1995). Interactive Spatial Data Analysis. Longman Scientific and Technical, Essex, England.

Bajaj, JS, Misra, A, Rajalakshmi, M, Madan, R (1993). Environmental Release of Chemicals and Reproductive Ecology. *Environmental Health Perspectives Supplements* 101(Suppl. 2): 125-130.

Baker, DB, Greenland, S, Mendlein, J, Harmon, P (1988). A health study of two communities near the Stringfellow waste disposal site. *Archives of Environmental Health* 43:325-334.

Band, P and Camus, M (1998). *Mortality study of Cape Breton County and Sydney, Nova Scotia. Standardized Comparisons with Canada, 1951 to 1994.* Environmental Health Directorate, Health Canada.

Birnbaum, LS (1995). Developmental effects of dioxins. *Environmental Health Perspectives* 103:89-94.

Ben Akiva, M, and Lerman, SR (1985). Discrete Choice Analysis: Theory and Application to Predict Travel Demand. The MIT Press, Cambridge MA.

Ben-Akiva, M. and D. Bolduc. Multinomial Probit with a Logit Kernel and a General Parametric Specificationof the Covariance Structure. Working paper, Massachusetts Institute of Technology, 1996.

Berry, M, and Bove, F (1997). Birth weight reductio- associated with residence near a hazardous waste landfill. *Environmental Health Perspectives* 105:856-861.

Beyea, J and Hatch, M (1999). Geogrpahic exposure modeling: a valuable extension of geographic information systems for use in environmental epidemiology. *Environmental Health Perspectives*, 107(Suppl. 1): 181-190.

Bhopal, RS, Tate, JA, Foy, C, Moffat, S, Phillimore, PR (1999). Residential proximity to industry and adverse birth outcomes. *The Lancet* 354:920-921

Bloom, AD, ed. (1981). *Guidelines for Studies of Human Populations Exposed to Mutagenic and Reproductive Hazards*. White Plains, New York: March of Dimes Birth Defects Foundation.

Briggs, D.J. (2000). Exposure Assessment, in Elliott, P. et al. (eds.) Spatial Epidemiology: Methods and Applications, Oxford, 335-359.

Buliung, RN and DeLuca, PF (2000). Spatial Patterns of Demand for Education: A Case Study. *Journal of Geographic Information and Decision Analysis*, 4(2): 37-51.

Burra, T, Jerrett, M, Burnett, RT and Anderson, M (2001). Conceptual and practical issues in the detection of local disease clusters: A study of mortality in Hamilton, Ontario. *Canadian Geographer* 46(2):160-171.

Burra, T (2002). *Reproductive and Psychosocial Health of Women Living in the Vicinity of the Tar Ponds, Sydney, NS.* Masters Thesis, School of Geography and Geology, McMaster University.

Burnett, RT, Dales, RE, Krewski, D, Vincent, R, Dann, T, Brook, JR (1995). Associations between ambient particulate sulfate admissions to Ontario hospitals for cardiac and respiratory diseases. *American Journal of Epidemiology*, 142:15-22.

Bushart, SP, Bush, B, Barnard, EL, Bott, A (1998). Volatilization of extensively dechlorinated polychlorinated biphenyls from historically contaminated sediments. *Environmental Toxicology and Chemistry*, 17:1927-1933.

Butler, WJ and Kalasinski, LA (1989). Statistical analysis of epidemiologic data of pregnancy outcomes. *Environmental Health Perspectives*, 79:223-227

Canadian British Consultants Limited (CBCL) and Conestoga-Rovers and Associates (CRA). 1999. *Phase 1 Site Assessment Muggah Creek Watershed Sydney, Nova Scotia.* Final Report No. 98628B prepared for Supply and Services Canada. Sydney, CBCL Limited and Conestoga Rovers and Associates.

CBRM (2003). "Cape Breton Regional Municipality – About CBRM". http://www.cbrm.ns.ca/portal/community/about/default.asp.

Chiarenzelli, J, Scrudato, R, Arnold, G, Wunderlich, M, Rafferty, D (1996). Volatilization of polychlorinated biphenyls from sediment during drying at ambient conditions. *Chemosphere*, 33:899-911.

Chiarenzelli, J, Bush, B, Casey, A, Barnard, E, Smith, B, O'Keefe, P et al. (2000). Defining sources of airborne polychlorinated biphenyls: evidence for the influence of microbially dechlorinated congeners from river sediment? Canadian Journal of Fisheries and Aquatic Sciences, 57:86-94.

Colborn, T, Vom Saal, FS, Soto, AM. Developmental effects of endocrine disrupting chemicals in wildlife and humans. *Environmental Health Perspectives*, 101:378-384.

Corporate Information Services of Devon County (2003). "The Townsend Material Deprivation Score". http://www.devon.gov.uk/dris/commstat/townsend.html.

Croen, LA, Shaw, GM, Sanbonmatsu, L, Selvin, S, Buffler, PA (1997). Maternal residential proximity to hazardous waste sites and risk of selected congenital malformations. *Epidemiology*, 8:347-354.

Cuzick, J (1998). Commentary: Clustering of anophthalmia and microphthalmia is not supported by the data. *British Medical Journal* 318:910-910.

De Wals, P (1999). Investigation of clusters of adverse reproductive outcomes, an overview. *European Journal of Epidemiology* 15:871-875.

Delfino, RJ, Murphy-Moulton, AM, Burnett, RT, Brook, JR, Becklake, MR (1997). Effects of air pollution on emergency room visits for respiratory illnesses in Montreal, Quebec. *American Journal of Respiratory and Critical Care Medicine*, 155: 568-576.

Diggle, PJ and Chetwynd, AG (1991). Second-order analysis of spatial clustering for inhomogenous populations. *Biometrics* 47:1155-1163.

Diggle and Morris (1996). Second-order analysis of spatial clustering. In: Methods for Investigating Localized Clustering of Disease. IARC Scientific Publications No. 135 (Eds. Alexander, FE and Boyle, P). Lyon, France

Dodds, L and Seviour, R (2001). Congenital anomalies and other birth outcomes among infants born to women living near a hazardous waste site in Sydney, Nova Scotia. *Canadian Journal of Public Health* 92:331-34.

Dolk, H, Vrijheid, M, Armstrong, B, Abramsky, L, Bianchi, F, Garne, E, Nelen, V, Robert, E, Scott, JES, Stone, D and Tenconi, R (1998a). Risk of congenital anomalies near hazardous-waste landfill sites in Europe: the Eurohazcon study. *The Lancet*, 352:423-427.

Dolk, H, Busby, A, Armstrong, BG, Walls, PH (1998b). Geographical variation in anophthalmia and microphthalmia in England, 1988-1994. *British Medical Journal*, 317:905-910.

Dolk, H (1999). The role of assessment of spatial variation and clustering in environmental surveillance of birth defects. *European Journal of Epidemiology* 15:839-845.

Dolk, H, Pattenden, S, Vrijheid, M, Thakrar, B, Armstrong, B (2000). Perinatal and Infant Mortality and Low Birth Weight among Residents near Cokeworks in Great Britain. *Archives of Environmental Health* 55(1):26-30.

Dummer, TJB, Dickinson, HO, Parker, L (2003). Adverse pregnancy outcomes around incinerators and crematoriums in Cumbria, north west England, 1956-93. *Journal of Epidemiology and Community Health*, 57:456-461.

Dunn, CE, Kingham, SP, Rowlingson, B, Bhopal, RS, Cockings, S, Foy, CJW, Acquilla, SD, Halpin, J, Diggle, P and Walker, D (2001). Analysing spatially referenced public health data: a comparison of three methodological approaches. *Health & Place* 7:1-12.

Elliott, P, Briggs, D, Morris, S, de Hoogh, C, Hurt, C, Kold Jensen, T, Maitland, I, Richardson, S, Wakefield, J, Jarup, L (2001). Risk of adverse birth outcomes in populations living near landfill sites. *British Medical Journal*, 323:363-368.

Elliott, SJ (1999). And the question shall determine the method. *Professional Geographer* 51:240-243.

Elliott, SJ, Eyles, J and DeLuca, P (2001). Mapping Health in the Great Lakes Areas of Concern: A User-Friendly Tool for Policy and Decision Makers. *Environmental Health Perspectives* 109(S6): 817 – 826.

English, PB, Kharrazi, M, Davies, S, Scalf, R, Waller, L and Neutra, R (2003). Changes in the spatial pattern of low birth weight in a southern California county: the role of individual and neighbourhood factors. *Social Science & Medicine* 56:2073-2088.

Environment Canada (1999). *Phase I Site Assessment: Muggah Creek Watershed, Sydney, Nova Scotia.* Final Report No. 98628B. Canadian British Consultants Limited (Sydney, NS) and Conestoga-Rovers & Associates (Waterloo, ON). 135 pp + appendices.

Evans, RG and Stoddart, GL (1990). Producing Health, Consuming Health Care. *Social Science and Medicine* 31:1347-1363.

Evans, RG (1994). Introduction. In Why are Some People Healthy and Others Not? The Determinants of Health of Populations. Eds. Evans, RG, Barer, ML and Marmor, TR. New York: Aldine De Gruyter.

Fein, GG, Jacobson, JL, Jacobson, SW, Schwartz, PM, Fowler, JK (1984). Prenatal exposure to polychlorinated biphenyls: effects on birth size and gestational age. *Journal of Pediatrics* 105:315-320.

Fellmann, JD, Getis, A, Getis, J (2001). Human Geography: Landscapes of Human Activities (6th Edition). McGraw-Hill, New York.

Fielder, HM, Poon-King, C, Palmer, SR, Moss, N, Coleman, G (2000). Assessment of impact on health of residents living near the Nant-y-Gwyddon landfill site: retrospective analysis. *British Medical Journal* 320:19-23.

Forand, SP, Talbot, TO, Druschel, C, Cross, PK (2002). Data quality and the spatial analysis of disease rates: congenital malformations in New York State. *Health & Place*, 8:191-199.

Forget, G and Lebel, J (2001). An ecosystem approach to human health. *International Journal of Occupational and Environmental Health* 7:S3-S38

Forssas, E, Gissler, M, Sihvonen, M, Hemminki, E (1999). Maternal predictors of perinatal mortality: the role of birthweight. *International Journal of Epidemiology*, 28:475-478.

Foster, WG, Jarrell, JF, Younglai, EV, Wade, MG, Arnold, DL, Jordan, S (1996). An overview of some reproductive toxicology studies conducted at Health Canada. *Toxicology and Industrial Health* 12(3/4):447-459.

Gatrell, AC (2002) Geographies of Health: An Introduction. London: Blackwell Publishers.

Gatrell, AC, Bailey, TC, Diggle, PJ and Rowlingson, BS (1996). Spatial point pattern analysis and its application in geographical epidemiology. *Transactions of the Institute of British Geographers* 21:256-274.

Gatrell, AC and Senior, ML (1999) Health and healthcare applications, pp 925-38 in Longley, P.A., Maguire, D.J., Goodchild, M.F. and Rhind, D.W. (eds) Geographical Information Systems, John Wiley, Chichester

Gazmararian, JA, Adams, MM, Pamuk, ER (1996). Associations between measures of socioeconomic status and maternal health behavior. *American Journal of Preventative Medicine* 12(2):108-15

Glei, DA and Goldman, N (2000). Understanding ethnic variation in pregnancy-related care in rural Guatemala. *Ethnicity and Health* 5(1):5-22.

Gochfeld, M (1995). Community exposure to hazardous waste. In Environmental Medicine, 635-646, eds Brooks,SM, Gochfeld, M, Herzstein, J, Jackson, RJ, Schenker, MB. St. Louis: Mosby.

Goldberg, MS, Goulet, L, Riberdy, H, Bonvalot, Y (1995). Low birth weight and premature births among infants born to women near a municipal solid waste landfill site in Montreal, Quebec. *Environmental Research* 69:37-50.

Goldman, L, Paigen, B, Magnant, MM, Highland, JH (1985). Low birth weight, prematurity and birth defects in children living near the hazardous waste site, Love Canal. *Hazardous Waste & Hazardous Materials* 2:209-23.

Gomez, J and Mattison, DR (1995). Female Reproductive System. In Environmental Medicine, 101-14, eds. Brooks, SM, Gochfeld, M, Herzstein, J, Jackson, RJ, Schenker, MB. St. Louis: Mosby.

Gouveia, N, Bremner, SA, Novaes, HMD (2004). Association between ambient air pollution and birth weight in Sao Paulo, Brazil. *Journal of Epidemiology and Community Health*, 58:11-17.

Guernsey, JR, Dewar, R, Weerasinghe, S, Kirkland, S and Veugelers, PJ (2000). Incidence of Cancer in Sydney and Cape Breton County, Nova Scotia, 1979-1997. *Canadian Journal of Public Health* 91(4): 285-292.

Ha, EH, Hong, YC, Lee, BE, Woo, BH, Schwartz, J, Christiani, DC (2001). Is Air Pollution a Risk Factor for Low Birth Weight in Seoul? *Epidemiology* 12(6):643-648.

Haining, RP (2003). Spatial Data Analysis: Theory and Practice. Cambridge University Press, London, UK.

Harvey, ER (1971). Sydney, Nova Scotia: An Urban Study. Clarke and Irwin, Toronto.

Health Canada (1999). *Toward a Healthy Future: The Second Report on the Health of Canadians*. Ottawa: Minister of Public Works and Government Services, Canada.

Health Canada (2000). *Canadian Perinatal Health Report*. Ottawa: Minister of Public Works and Government Services, Canada

Health Canada (2003). *Canadian Perinatal Health Report 2003*. Ottawa: Minister of Public Works and Government Services, Canada

Hertz-Piccioto, I, Pastore, LM, Beaumont, JJ (1996). Timing and patterns of exposures during pregnanct and their implications for study methods. *American Journal of Epidemiology*, 143: 597-607.

Higgs, G, Gould, M (2001). Is there a role for GIS in the 'new NHS'? *Health & Place*, 7(3):247-259.

Holmes, LB (1999). Need for inclusion and exclusion criteria for the structural abnormalities recorded in children born from exposed pregnancies. *Teratology*, 59:1-2.

JAG (2003). "Sydney Tar Ponds Coke Ovens Cleanup." http://www.muggah.org (25 Oct 2003).

JDAC Environment (2002). *Phase III Environmental Site and Risk Assessments Coke Ovens Site: Phase II/III Environmental Site Assessment: Muggah Creek Watershed.* Volume 1: Sections 1-9, October, 2002.

Joseph, KS and Kramer, MS (1997). Recent trends in infant mortality rates and proportions of low birth weight live births in Canada. *Canadian Medical Association Journal* 157:535-541.

Kammann, EE and Wand, MP (2002). Geoadditive models. *Applied Statistics*, 52(1):1-18.

Kharrazi, M, VonBehren, J, Smith, M, Lomas, T, Armstrong, M, Broadwin, R, Blake, E, McLaughlin, B, Worstell, G, Goldman, L (1997). A community based study of adverse pregnancy outcomes near a large hazardous waste landfill in California. *Toxicology and Industrial Health*, 13:299-310.

Kramer, MS (1987a). Determinants of low birth weight: methodological assessment and meta-analysis. *Bulletin of the World Health Organization* 65:663-737.

Kramer, MS (1987b). Intrauterine growth and gestational duration determinants. *Pediatrics* 80:502-511.

Kramer, MS, Olivier, M, McLean, FH, Dougherty, GE, Willis, DM, Usher, RH (1990). Determinants of fetal growth and body proportionality. *Pediatrics* 86:18-26.

Kramer, MS, Platt, R, Yang, H, Joseph, KS, Wen, SW, Morin, L, Usher, RH (1998). Secular Trends in Preterm Birth. *Journal of the American Medical Association*, 280(21):1849-1854.

Kramer, MS, Goulet, L, Lydon, J, Seguin, L *et al.* (2001). Socio-economic disparities in preterm birth: causal pathways and mechanisms. *Paediatric and Perinatal Epidemiology*, 15(Suppl. 2): 104-123.

Kramer, MS, Liu, S, Luo,Z, Yuan, H, Platt, RW, Joseph, KS (2002). Analysis of perinatal mortality and its components: time for a change? *American Journal of Epidemiology*, 156(6):493-497.

Kramer, MS (2003). The epidemiology of adverse pregnancy outcomes: an overview. *The Journal of Nutrition* 133:1592S-1596S.

Landgren, O (1996). Environmental pollution and delivery outcome in southern Sweden: a study with central registries. *Acta Paediatrica*, 85:1361-1364.

Leon, DA (1991). Influence of birth weight on differences in infant mortality by social class and legitimacy. *British Medical Journal* 303:964-967)

Levine, N (2002). CrimeStat: A Spatial Statistics Program for the Analysis of Crime Incident Locations (v.2.0). Ned Levine & Associates, Houston, TX, and the National Institute of Justice, Washington DC.

Lin, MC, Yu, HS, Tsai, SS, Cheng, BH, Hsu, TY, Wu, TN and Yang, CY (2001). Adverse Pregnancy Outcome in a Petrochemical Polluted Area in Taiwan. *Journal of Toxicology and Environmental Health, Part A.* 63:565-574.

Lindbohm, ML, Taskinen, H, Kyyronen, P, Sallmen, M, Antilla, A, Hemminki, K (1992). Effects of prenatal occupational exposure to solvents and lead on spontaneous abortion. *Scandanavian Journal of Work, Environment and Health* 18:37-39.

Mackay, D and Webster, E (1998). Linking emissions to prevailing concentrations – exposure on a local scale. *Environmetrics* 9:541-553

MacMillen, DP (1992). Probit with spatial autocorrelation. *Journal of Regional Science*, 32(3):335-348.

Malachowski, MJ (1995). Health Effects of Toxic Substances. Government Institutes Inc. Rockville, Maryland.

Mann, CJ (2003). Observational research methods. Research design II: cohort, cross sectional and case-control studies. *Emergency Medicine Journal*, 20:54-60

Mao, Y, Morrison, H and Semenciew, R (1985). *Mortality in Cape Breton, Nova Scotia, 1971-1983*. Special Report No. 11. Chronic Diseases in Canada. Bureau of Non-Communicable Disease Division, Bureau of Epidemiology, Laboratory Centre for Disease Control, Health and Welfare Canada.

Marshall, EG, Gensburg, LJ, Deres, DA, Geary, NS, Cayo, MR (1993). Maternal residential exposure to hazardous wastes and risk of central nervous system and musculoskeletal birth defects. *Archives of Environmental Health*, 52:416-425.

Mattison, DR (1983). The mechanisms of reproductive toxins. *American Journal of Industrial Medicine*, 4:65-79

Mersereau, H, Dugandzic, R, Campbell, R and MacIntyre, P (1999). "*Health Risk Perceptions of Sydney Residents: Risk Communication of Sydney Tar Ponds Information.*" Technical Report, UCCB Press. 33+vi pages, plus appendices.

Michal, F, Grigor, KM, Negro-Vilar, A, Skakkebaek, NE (1993). Impact of the environment on reproductive health: executive summary. *Environmental Health Perspectives Supplements*, 101(Suppl. 2): 159-167.

Mohammadian, A, Kanaroglou, P and Haider, M (2003). *Spatial Multinomial Logit Model: A Land Development Choice Model Considering Spatial Dependencies*. 82nd Annual Transportation Research Board Meeting, Washington DC. January 2003.

Moutquin, JM (2003). Classification and heterogeneity of preterm birth. *British Journal of Obstetrics and Gynaecology*, 110(Suppl. 20):30-33.

Myers, GJ, Marsh, DO, Davidson, PW, Cox, X, Shamlaye, CF, Tanner, H (1995). Main neurodevelopmental study of Seychellois children following in utero exposure to Methylmercury from a maternal fish diet: outcome at six months. *Neurotoxicology* 16:653-664.

Negro-Vilar, A (1993). Stress and other environmental factors affecting fertility in and women: overview. *Environmental Health Perspectives Supplements*, 101(Suppl. 2): 59-64.

Nova Scotia Department of Health and the Cape Breton District Health Authority (2001). *Lead and Arsenic Biological Testing Program in Residential Areas Near the Coke Ovens Site*. http://www.muggah.org/site/projects/archives/10.pdf. Accessed on September 22, 2004.

NSRCP (2000). *1999 Nova Scotia Perinatal Database Report*. Halifax: Reproductive Care Program of Nova Scotia.

Pattenden, S, Dolk, H, Vrijheid, M (1999). Inequalities in low birthweight: parental social class, area deprivation and "lone mother" status. *Journal of Epidemiology and Community Health* 53:355-358.

Perera, FP, Rauh, V, Tsai, WY, Kinney, P, Camann, D, Barr, D, Bernart, T, Garfinkel, R, Tu, YH, Diaz, D, Dietrich, J, Whyatt, RM (2003). Effects of transplacental exposure to environmental pollutants on birth outcomes in a multiethnic population. *Environmental Health Perspectives*, 111(2):201-205

Private Communication from John Fahey: Nova Scotia Reproductive Care Program, IWK Grace Hospital, Halifax, Nova Scotia, March 2003.

Private Communication from Dr. Bruce Wainman: Department of Obstetrics and Gynecology, McMaster University, Hamilton, Ontario, April 2003.

Rainham, D (2002). Risk communication and public response to industrial chemical contamination in Sydney, Nova Scotia: A case study. *Journal of Environmental Health*, 65(5): 26-31.

Reader, S (2000). Using survival analysis to study spatial point patterns in geographical epidemiology. *Social Science & Medicine* 50(7-8):985-1000.

Reader, S (2001). Detecting and analyzing clusters of low-birth weight incidence using exploratory spatial data analysis. *GeoJournal* 53:149-159.

Robson, S, Chan, A, Keane, RJ, Luke, CG (2001). Subsequent birth outcomes after an unexplained stillbirth: preliminary population based retrospective cohort study. *Australian and New Zealand Journal of Obstetrics and Gynecology*, 41:29-35.

Rogers, JF, Thompson, SJ, Addy, CL, McKeown, RE, Cowen, DJ, Decoufle, P (2000). Association of very low birth weight with exposures to environmental sulfur dioxide and total suspended particulates. *American Journal of Epidemiology* 151:602-613.

Rowlingson, BS and Diggle, PJ (1993). SPLANCS: Spatial point pattern analysis code in S-PLUS. *Computers and Geosciences* 19(5):627-655.

Rushton, G and Lolonis, P (1996). Exploratory spatial analysis of birth defect rates in an urban population. *Statistics in Medicine* 15:717-726.

Rushton, G (2003). Public health, GIS and spatial analytic tools. *Annual Review of Public Health* 24:43-56.

Rylander, L, Stromberg, U, Hagmar, L (2000). Lowered birth weight among infants born to women with a high intake of fish contaminated with persistent organochlorine compounds. *Chemosphere*, 43:895-901

Savitz, DA and Harlow SD (1991). Selection of reproductive health end points for environmental risk assessment. *Environmental Health Perspectives*, 90:159-164.

Scott, KE and Usher, R (1966). Fetal malnutrition: its incidence, causes and effects. American Journal of Obstetrics and Gynecology, 94(7): 951-963.

Sharara, FI, Seifer, DB, Flaws, JA (1998). Environmental toxicants and female reproduction. *Fertility and Sterility* 70(4): 613-622.

Smith, GD, Whitley, E, Dorling, D, Gunnell, D (2001). Area based measures of social and economic circumstances: cause specific mortality patterns depend on the choice of index. *Journal of Epidemiology and Community Health* 55:149-150.

Sosniak, WA, Kaye, WE, Gomez, TM (1994). Data linkage to explore the risk of low birthweight associated with maternal proximity to hazardous waste sites from the National Priorities List. *Archives of Environmental Health*, 52:416-425.

Statistics Canada (1996a). Community Profile of Cape Breton Regional Municipality. http://www.statcan.ca. Accessed on June 27, 2004.

Statistics Canada (1996b). 1996 Census. Ottawa, Ontario.

Statistics Canada (1996c). Postal Code Conversion File Reference Guide. Ottawa, Ontario.

Stephansson, O, Dickman, PW, Johansson, ALV, Cnattingius, S (2000). Maternal hemoglobin concentration and risk of stillbirth. Journal of the American Medical Association, 284(20):2611-2617.

Stephansson O, Dickman, PW, Johansson, ALV, Cnattingius, S (2001). The influence of socioeconomic status on stillbirth risk in Sweden. *International Journal of Epidemiology*, 30:1296-1301.

Sullivan, FM (1993). Impact of the Environment on Reproduction from Conception to Parturition. *Environmental Health Perspectives Supplements*. 101(Suppl. 2): 13-18.

Surkan, PJ, Stephansson, O, Dickman, PW, Cnattingius, S (2004). Previous preterm and small-gestational age births and the subsequent risk of stillbirth. *New England Journal of Medicine*, 350(8):777-785.

Taskinen, HC (1993). Epidemiological studies in monitoring reproductive effects. *Environmental Health Perspectives*, 101(Suppl 3): 279-283.

Tough, SC, Svenson, LW, Johnston, DW, Schopflocher, D (2001). Characteristics of preterm delivery and low birthweight among 113,994 infants in Alberta: 1994-1996. *Canadian Journal of Public Health*, 92(4):276-280.

Train, K. (1993). Qualitative Choice Analysis: Theory, Econometrics and an Application to Automobile Demand. The MIT Press, Cambridge, MA.

TRB (2001). Discrete Dependent Variable Models. Chapter 5. http://gulliver.trb.org/publications/nchrp/cd-22/v2chapter5.html.

Vanden Heuvel, J.P., Clark, G.C., Kohn, M.C., Tritscher, A.M., Greenlee, W.F., Lucier, G.Wand Bell, D.A. (1994). Dioxin-responsive genes: examination of dose-response relationships using quantitative reverse transcriptase-polymerase chain reaction. *Cancer Research*, 54:62-68.

Vassilev, ZP, Robson, MG, Klotz, JB (2001). Associations of polycyclic organic matter in outdoor air with decreased birth weight: a pilot cross-sectional analysis. Journal of Toxicology and Environmental Health, Part A, **64**:595-605.

Veugelers, PJ and Guernsey, JR (1999a). Health deficiencies in Cape Breton County, Nova Scotia, Canada, 1950-1995. *Epidemiology* 10(5):495-499

Veugelers, PJ and Guernsey, JR (1999b). Sensitivity analysis of selective migration in ecological comparisons of health. *Epidemiology* 10(6):784-785

Vrijheid, M (2000). Health effects of residence near hazardous waste landfill sites: a review of epidemiologic literature. *Environmental Health Perspectives*, 108(Suppl 1): 101-112

Vrijheid, M, Dolk, H, Stone, D, Abramsky, L, Alberman, E, Scott, JES (2000). Socioeconomic inequalities in risk of congenital anomaly. *Archives of Disease in Childhood* 82:349-352.

Vrijheid, M, Dolk, H, Armstrong, B, Boschi, G, Busby, A, Jorgensen, T, Pointer, P and the EUROHAZCON collaborative group (2002a). Hazard potential ranking of hazardous waste landfill sites and risk of congenital anomalies. *Journal of Occupational and Environmental Medicine*, 59:768-776.

Vrijheid, M, Dolk, H, Armstrong, B, Abramsky, L, Bianchi, F, Fazarinc, I, Garne, E, Ide, R, Nelen, V, Robert, E, Scott, JES, Stone, D, Tenconi, R (2002b). Chromosomal congenital anomalies and residence near hazardous waste landfill sites. *The Lancet*, 359:320-322.

Walter, SD (1991). The ecologic method in the study of environmental health I & II. *Environmental Health Perspectives*, 94, 61-73.

Watier, L, Richardson, S and Hemon, D (1997). Accounting for pregnancy dependence in epidemiologic studies of reproductive outcomes. *Epidemiology*, 8:629-36

Weinberg, CR and Wilcox, AJ (1998). Reproductive Epidemiology. In: Rothman, KJ, Greenland, S (eds.) Modern Epidemiology. (2nd Edition) Lippincott, Williams & Wilkins. Chapter 29, pp 585-608.

WHO (1948). Preamble to the Constitution of the World Health Organization as adopted by the International Health Conference, New York, 19-22 June, 1946; signed on 22 July 1946 by the representatives of 61 States (Official Records of the World Health Organization, no. 2, p. 100) and entered into force on 7 April 1948.

WHO (1986). *Evaluation of the Global Strategy for Health for All by the Year 2000.* Geneva: World Health Organization.

Wrigley, N (1985). Categorical Data Analysis for Geographers and Environmental Scientists. Blackburn Press, New Jersey.

Yang, CY, Chiu, HF, Tsai, SS, Chang, CC and Chuang, HY (2002). Increased risk of preterm delivery in areas with cancer mortality problems from petrochemical complexes. *Environmental Research A* 89:195-200.

APPENDIX I

LIST OF COMMONLY USED ABBREVIATIONS

ATSDR – Agency for Toxic Substances and Disease Regstry
CBRM – Cape Breton Regional Municipality
COR – Conditional Odds Ratio
CSR – Complete Spatial Randomness
DMTI – Desktop Mapping Technologies Limited
ESDA – Exploratory Spatial Data Analysis
GIS – Geographic Information Systems
HSWG – Health Studies Working Group
ICB – Industrial Cape Breton
JAG – Joint Action Group
LBW – Low birth weight
MNL – Multinomial Logistic regression model
OR – Odds Ratio
PAH – Polyaromatic Hydrocarbons
PCB – Polychlorinated Biphenyls
PCCF – Postal Code Conversion File
RCP – Reproductive Care Program
SES – Socio – economic status
SYSCO – Sydney Steel Corporation
VOC – Volatile Organic Compounds
WHO – World Health Organization

APPENDIX II

List of Variables Obtained From the Atlee Perinatal Database

Nova Scotia Atlee Perinatal Database
Reproductive Care Program of Nova Scotia
[902)470-6798]
RCP@iwk.NSHealth.Ca

Data Extraction for Project 4036
GIS & associated variables for Tar Ponds/Coke Ovens Impact on Reproductive Health

-----Variables Ordered by Position-----

#	Variable	Label
1	ACONTACT_ ID	Unique Identifier
2	LATITUDE	Latitude
3	LONGITUDE	Longitude
4	MUN_CODE	Municipality Code
5	POST_CODE	Postal Code
6	LL_Base	Basis for Latitude and Longitude
7	BIRTHID	Birth ID
8	IVINS	Indicator: Velamentous Insertion of Cord
9	IVINSC	Modifier: Velamentous Insertion of Cord
10	IMINS	Indicator: Marginal Insertion of Cord
11	IMINSC	Modifier: Marginal Insertion of Cord
12	ICIRV	Indicator: Circumvallate Placenta
13	ICIRVC	Modifier: Circumvallate Placenta
14	IANOD	Indicator: Amnionodosum
15	IANODC	Modifier: Amnionodosum
16	ISUCL	Indicator: Succenturiate Lobe
17	ISUCLC	Modifier: Succenturiate Lobe
18	ICHOR	Indicator: Chorioamnionitis, Marked or Severe
19	ICHORC	Modifier: Chorioamnionitis, Marked or Severe
20	IFUNI	Indicator: Funisitis
21	IFUNIC	Modifier: Funisitis
22	ISUMA	Indicator: Single Umbilical Artery
23	ISUMAC	Modifier: Single Umbilical Artery
24	ICPRO	Indicator: Cord Prolapse
25	ICPROC	Modifier: Cord Prolapse
26	IPLUC	Indicator: Miscellaneous Placenta and/or Umbilical Cord Abnormality
27	IPLUCC	Modifier: Miscellaneous Placenta and/or Umbilical Cord Abnormality
28	IWAST	Indicator: Fetal Clinical Soft Tissue Wasting
29	IWASTC	Modifier: Fetal Clinical Soft Tissue Wasting

30	ICARD	Indicator: Congenital Heart Disease
31	ICARDC	Modifier: Congenital Heart Disease
32	IDUCT	Indicator: Ductus Syndrome of Prematurity
33	IDUCTC	Modifier: Ductus Syndrome of Prematurity
34	ICARR	Indicator: Cardiac Arrhythmia
35	ICARRC	Modifier: Cardiac Arrhythmia
36	IDEXT	Indicator: Dextrocardia
37	IDEXTC	Modifier: Dextrocardia
38	IPFCS	Indicator: Persistent Fetal Circulation Syndrome
39	IPFCSC	Modifier: Persistent Fetal Circulation Syndrome
40	IPABS	Indicator: Parabiotic Syndrome
41	IPABSC	Modifier: Parabiotic Syndrome
42	ICVSA	Indicator: Miscellaneous Cardiovascular Anomaly
43	ICVSAC	Modifier: Miscellaneous Cardiovascular Anomaly
44	IIINA	Indicator: Intestinal Atresia
45	IIINAC	Modifier: Intestinal Atresia
46	IIINS	Indicator: Intrinsic Intestinal Stenosis
47	IIINSC	Modifier: Intrinsic Intestinal Stenosis
48	IEINO	Indicator: Extrinsic Intestinal Obstruction
49	IEINOC	Modifier: Extrinsic Intestinal Obstruction
50	ITEFA	Indicator: Tracheo-Esophageal Fistula/Atresia
51	ITEFAC	Modifier: Tracheo-Esophageal Fistula/Atresia
52	IMALR	Indicator: Intestinal Malrotation
53	IMALRC	Modifier: Intestinal Malrotation
54	IBILA	Indicator: Biliary Atresia
55	IBILAC	Modifier: Biliary Atresia
56	IHIRD	Indicator: Hirschsprung's Disease
57	IHIRDC	Modifier: Hirschsprung's Disease
58	IIMPA	Indicator: Imperforate Anus
59	IIMPAC	Modifier: Imperforate Anus
60	IGIAN	Indicator: Miscellaneous GI Anomaly
61	IGIANC	Modifier: Miscellaneous GI Anomaly
62	IVOLV	Indicator: Volvulus
63	IVOLVC	Modifier: Volvulus
64	IPHYP	Indicator: Pulmonary Hypoplasia/Agenesis
65	IPHYPC	Modifier: Pulmonary Hypoplasia/Agenesis
66	IDIAH	Indicator: Diaphragmatic Hernia
67	IDIAHC	Modifier: Diaphragmatic Hernia
68	BRGC	Indicator: Bronchogenic Cyst
69	IBRGCC	Modifier: Bronchogenic Cyst
70	IRAUN	Indicator: Miscellaneous Respiratory Anomaly
71	IRAUNC	Modifier: Miscellaneous Respiratory Anomaly
72	IHYPD	Indicator: Hypoplasia of Diaphragm
73	IHYPDC	Modifier: Hypoplasia of Diaphragm

74	ICTLP	Indicator: Cleft Lip and/or Palate
75	ICTLPC	Modifier: Cleft Lip and/or Palate
76	IBCSF	Indicator: Branchial Cleft Anomaly
77	IBCSFC	Modifier: Branchial Cleft Anomaly
78	ICTAR	Indicator: Cataracts
79	ICTARC	Modifier: Cataracts
80	ISKTP	Indicator: Pre-auricular Skin Tag, Pit or Sinus
81	ISKTPC	Modifier: Pre-auricular Skin Tag, Pit or Sinus
82	IEARS	Indicator: Stenosis or Atresia of External Auditory Meatus or Canal
83	IEARSC	Modifier: Stenosis or Atresia of External Auditory Meatus or Canal
84	ICOPA	Indicator: Opacities of Cornea, Congenital
85	ICOPAC	Modifier: Opacities of Cornea, Congenital
86	IMCGN	Indicator: Micrognathia
87	IMCGNC	Modifier: Micrognathia
88	IMCOP	Indicator: Microphthalmia
89	IMCOPC	Modifier: Microphthalmia
90	IGLAC	Indicator: Glaucoma
91	IGLACC	Modifier: Glaucoma
92	ICOLB	Indicator: Coloboma
93	ICOLBC	Modifier: Coloboma
94	IINLD	Indicator: Congenital Impatency of the Naso-Lacrimal Duct
95	IINLDC	Modifier: Congenital Impatency of the Naso-Lacrimal Duct
96	IEAUC	Indicator: Miscellaneous Eye, Ear, Nose, Mouth, Throat Anomaly
97	IEAUCC	Modifier: Miscellaneous Eye, Ear, Nose, Mouth, Throat Anomaly
98	ICHOA	Indicator: Choanal Atresia
99	ICHOAC	Modifier: Choanal Atresia
100	ITHYG	Indicator: Thyroglossal Anomalies
101	ITHYGC	Modifier: Thyroglossal Anomalies
102	ICORC	Indicator: Cryptorchidism
103	ICORCC	Modifier: Cryptorchidism
104	IINGH	Indicator: Inguinal Hernia
105	IINGHC	Modifier: Inguinal Hernia
106	IFEMH	Indicator: Femoral Hernia
107	IFEMHC	Modifier: Femoral Hernia
108	IHSPD	Indicator: Hypospadias Complex
109	IHSPDC	Modifier: Hypospadias Complex
110	IEPSD	Indicator: Epidpadias
111	IEPSDC	Modifier: Epidpadias

112	IPOLY	Indicator: Polycystic Kidney
113	IPOLYC	Modifier: Polycystic Kidney
114	IAGEN	Indicator: Agenesis/Hypoplasia/Atrophy of Kidney
115	IAGENC	Modifier: Agenesis/Hypoplasia/Atrophy of Kidney
116	IHNHU	Indicator: Hydronephrosis and/or Hydroureter and/or Renal Pelviectasis
117	IHNHUC	Modifier: Hydronephrosis and/or Hydroureter and/or Renal Pelviectasis
118	IRDYS	Indicator: Renal Dysplasia
119	IRDYSC	Modifier: Renal Dysplasia
120	IDSYS	Indicator: Double Urinary System
121	IDSYSC	Modifier: Double Urinary System
122	IGAGN	Indicator: Genital Agenesis/Hypoplasia
123	IGAGNC	Modifier: Genital Agenesis/Hypoplasia
124	IRECT	Indicator: Renal Ectopia
125	IRECTC	Modifier: Renal Ectopia
126	IUROB	Indicator: Urinary Obstruction
127	IUROBC	Modifier: Urinary Obstruction
128	IIHYM	Indicator: Imperforate Hymen
129	IIHYMC	Modifier: Imperforate Hymen
130	IUBAN	Indicator: Urinary Bladder Anomalies
131	IUBANC	Modifier: Urinary Bladder Anomalies
132	IGUAN	Indicator: Miscellaneous G.U. Anomaly
133	IGUANC	Modifier: Miscellaneous G.U. Anomaly
134	IHANG	Indicator: Hemangioma
135	IHANGC	Modifier: Hemangioma
136	IPIGN	Indicator: Pigmented Nevus
137	IPIGNC	Modifier: Pigmented Nevus
138	IDPIG	Indicator: Depigmented Skin Lesions
139	IDPIGC	Modifier: Depigmented Skin Lesions
140	ICDER	Indicator: Congenital Dermatosis
141	ICDERC	Modifier: Congenital Dermatosis
142	ISUPN	Indicator: Supernumerary Nipple
143	ISUPNC	Modifier: Supernumerary Nipple
144	IABBR	Indicator: Absent Breasts
145	IABBRC	Modifier: Absent Breasts
146	IAMNB	Indicator: Deformities Due to Amniotic Bands, Amniotic Band Syndrome, Early Amnion Rupture Sequence
147	IAMNBC	Modifier: Deformities Due to Amniotic Bands, Amniotic Band Syndrome, Early Amnion Rupture Sequence

148	IICTH	Indicator: Ichthyosis
149	IICTHC	Modifier: Ichthyosis
150	ICUTL	Indicator: Cutis Laxa/Hyperelastica
151	ICUTLC	Modifier: Cutis Laxa/Hyperelastica
152	ISKAN	Indicator: Miscellaneous Skin Abnormality
153	ISKANC	Modifier: Miscellaneous Skin Abnormality
154	INECK	Indicator: Short Neck Disorders
155	INECKC	Modifier: Short Neck Disorders
156	ICLFT	Indicator: Club Foot
157	ICLFTC	Modifier: Club Foot
158	ICHIP	Indicator: Congenital Hip
159	ICHIPC	Modifier: Congenital Hip
160	IPOLD	Indicator: Polydactyly
161	IPOLDC	Modifier: Polydactyly
162	ISYND	Indicator: Syndactyly
163	ISYNDC	Modifier: Syndactyly
164	ICSTN	Indicator: Craniosynostosis/Craniostenosis
165	ICSTNC	Modifier: Craniosynostosis/Craniostenosis
166	IUMBD	Indicator: Umbilical Defect
167	IUMBDC	Modifier: Umbilical Defect
168	IHAND	Indicator: Other Anomalies of the Hand/Foot
169	IHANDC	Modifier: Other Anomalies of the Hand/Foot
170	IPHOC	Indicator: Phocomelia/Amelia/Limb Reduction
171	IPHOCC	Modifier: Phocomelia/Amelia/Limb Reduction
172	IACON	Indicator: Osteochondroplasia
173	IACONC	Modifier: Osteochondroplasia
174	ITORT	Indicator: Torticollis
175	ITORTC	Modifier: Torticollis
176	IVERT	Indicator: Vertebral anomalies
177	IVERTC	Modifier: Vertebral anomalies
178	IAPCM	Indicator: Absence/hypoplasia of pectoralis major
179	IAPCMC	Modifier: Absence/hypoplasia of pectoralis major
180	ISMEL	Indicator: Sirenomelus
181	ISMELC	Modifier: Sirenomelus
182	IBBDF	Indicator: Osteogenesis Imperfecta
183	IBBDFC	Modifier: Osteogenesis Imperfecta
184	ITHDW	Indicator: Thanatophoric Dwarfism
185	ITHDWC	Modifier: Thanatophoric Dwarfism
186	IRADA	Indicator: Radial Aplasia/Hypoplasia
187	IRADAC	Indicator: Miscellaneous anomaly of musculo-skeletal system
189	IMSANC	Modifier: Miscellaneous anomaly of musculo-skeletal system
190	IMYOP	Indicator: Myopathy

191	IMYOPC	Modifier: Myopathy
192	ICNSA	Indicator: C.N.S. Anomaly, Miscellaneous
193	ICNSAC	Modifier: C.N.S. Anomaly, Miscellaneous
194	IMICC	Indicator: Microcephaly
195	IMICCC	Modifier: Microcephaly
196	ISBIF	Indicator: Incomplete Closure of Neural Tube
197	ISBIFC	Modifier: Incomplete Closure of Neural Tube
198	IANEN	Indicator: Anencephaly
199	IANENC	Modifier: Anencephaly
200	IHYDN	Indicator: Hydranencephaly
201	IHYDNC	Modifier: Hydranencephaly
202	ICHYP	Indicator: Brain Hypoplasia
203	ICHYPC	Modifier: Brain Hypoplasia
204	IMEGC	Indicator: Megalocephaly
205	IMEGCC	Modifier: Megalocephaly
206	IDIAS	Indicator: Diastematomyelia
207	IDIASC	Modifier: Diastematomyelia
208	ICLOP	Indicator: Cyclopia-Arhinencephaly Series
209	ICLOPC	Modifier: Cyclopia-Arhinencephaly Series
210	IARMC	Indicator: Arthrogryposis/contractures
211	IARMCC	Modifier: Arthrogryposis/contractures
212	IPHAK	Indicator: Phakomatoses
213	IPHAKC	Modifier: Phakomatoses
214	ISPMA	Indicator: Spinal Muscular Atrophy
215	ISPMAC	Modifier: Spinal Muscular Atrophy
216	ICHRM	Indicator: Multiple Anomalies Due to Chromosomal Aberrations
217	ICHRMC	Modifier: Multiple Anomalies Due to Chromosomal Aberrations
218	IMDEL	Indicator: Chromosomal Deletion
219	IMDELC	Modifier: Chromosomal Deletion
220	INOTC	Indicator: Multiple Anomalies Not Due to Chromosomal Aberrations
221	INOTCC	Modifier: Multiple Anomalies Not Due to Chromosomal Aberrations
222	IOLIG	Indicator: Oligohydramnios Syndrome
223	IOLIGC	Modifier: Oligohydramnios Syndrome
224	IIKDM	Indicator: Infant of diabetic mother
225	MajrCard	Number of major cardiac anomalies
226	MajrCNvS	Number of major central nervous system anomalies
227	MajrEENT	Number of major ear, eye, nose & throat anomalies
228	MajrGenU	Number of major genito-urinary anomalies
229	MajrGInt	Number of major gastro-intestinal anomalies
230	MajrICan	Number of major inguinal canal anomalies

231	MajrMSkl	Number of major musculo-skeletal anomalies
232	MajrMult	Number of major multiple anomalies
233	MajrResp	Number of major respiratory anomalies
234	MajrSkin	Number of major skin anomalies
235	MajrMeta	Number of major metabolic anomalies
236	MinrCard	Number of minor cardiac anomalies
237	MinrCNvS	Number of minor central nervous system anomalies
238	MinrEENT	Number of minor ear, eye, nose & throat anomalies
239	MinrGenU	Number of minor genito-urinary anomalies
240	MinrGInt	Number of minor gastro-intestinal anomalies
241	MinrICan	Number of minor inguinal canal anomalies
242	MinrMSkl	Number of minor musculo-skeletal anomalies
243	MinrMult	Number of minor multiple anomalies
244	MinrResp	Number of minor respiratory anomalies
245	MinrSkin	Number of minor skin anomalies
246	MinrMeta	Number of minor metabolic anomalies
247	MajrAnom	Number of major anomalies
248	MinrAnom	Number of minor anomalies
249	NEAdmsDT	Admission SAS Datetime for Transfer/Re-admission
250	NEDethDT	Death SAS Datetime
251	NEPRMCOD	Neonate's Primary Cause of Death
252	TIMNGOFD	Timing of Foetal Death
253	BTPRMCOD	Infant's Primary Cause of Death
254	BTDethDT	Death SAS Datetime
255	DLPSTCOD	Postal Code
256	DLMUNCOD	Municipality Code
257	DLMRSTAT	Marital Status
258	DLGRAVID	# of Pregnancies, Including the Present One
259	DLPARA	# of Pregnancies, Excluding the Present, with >= 500g Birth
260	DLABORTS	# of Pregnancies, Excl. the Present, with Non-viable Foetus
261	DLPRVFTD	# Previous Deaths of Foetus >= 500g
262	DLPRVNND	# Previous Deaths of Neonate >= 500g
263	DLPRVLBW	# Previous Births < 2500 g
264	DLPRESMK	# Cigarettes / Day Pre-pregnancy
265	DLVS1SMK	# Cigarettes / Day @ 1st Pre-natal Visit
266	DLADMREA	Reason for Delivery Admission
267	DLNUMFET	# of Foetuses
268	ADMITSMK	# Cigarettes / Day @ Admission
269	DLAdmsDT	Admission SAS Datetime
270	MOTHERID	Mother's Person ID
271	DMGESAGE	Gestational Age Estimate (combined) (lowest)

		(weeks)
272	DMMATAGE	Mother's Age
273	MPHIP	Indicator: Hypertensive Disease in Previous Pregnancy
274	MPECL	Indicator: Previous Eclampsia
275	MPLOH	Indicator: Previous Anemia
276	MPGLD	Indicator: Previous Gestational Diabetes
277	MABRT	Indicator: Abortions
278	MCHTD	Indicator: Chronic Hypertensive Disease
279	MPIHT	Indicator: Pregnancy-Induced Hypertension
280	MECLP	Indicator: Eclampsia
281	MHEMS	Indicator: Hypermesis Gravidarum
282	MFGAS	Indicator: Fetal Growth Concerns
283	MPOST	Indicator: Post Dates
284	MSUFA	Indicator: Suspected Fetal Anomaly
285	MDRUG	Indicator: Maternal Drug Use During Present Pregnancy and/or Envirnonmental Exposure
286	MABUS	Indicator: Chemical Abuse
287	MJAUN	Indicator: Jaundice
288	MCRIN	Indicator: Endocrine
289	MANEM	Indicator: Anemia
290	MMATD	Indicator: Maternal Death
291	MFDTH	Indicator: Fetal Death
292	MPIHTC	Modifier: Pregnancy-Induced Hypertension
293	Birth_Weight	Rounded down to 500 g
294	Outcome	Pregnancy outcome (longest follow-up)

APPENDIX III

OUTCOME OF PREGNANCY ON LAST FOLLOW-UP AS CODED IN THE ATLEE PERINATAL DATABASE

Table 1: Outcome of pregnancy on last follow-up

Field Label	Field Definition	Number of Observations
FTD	Fetal Death	157
END	Early Neonatal Death (0-7 days)	103
LND	Late Neonatal Death (7-28 days)	16
NND	Neonatal Death (0-28 days)	119
IND	Post-Neonatal Death (28-365 days)	16
PID	Post-Infant Death (> 365 days)	0
LVD	Lived	21797

APPENDIX IV

Bivariate K-Functions

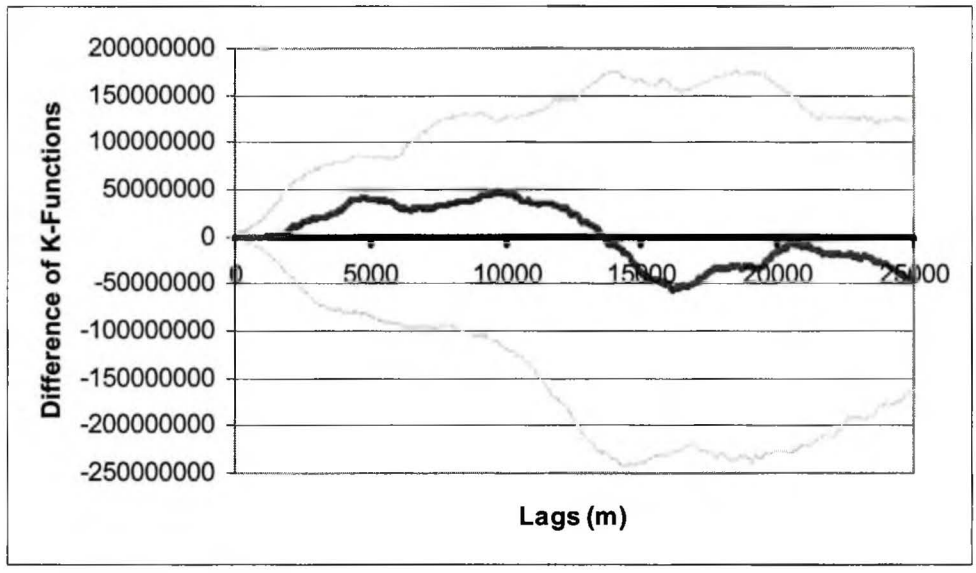


Figure A4.1. Difference of K-Functions for Congenital Anomalies, 1988 – 2002 using the entire set of normal births for the control population.

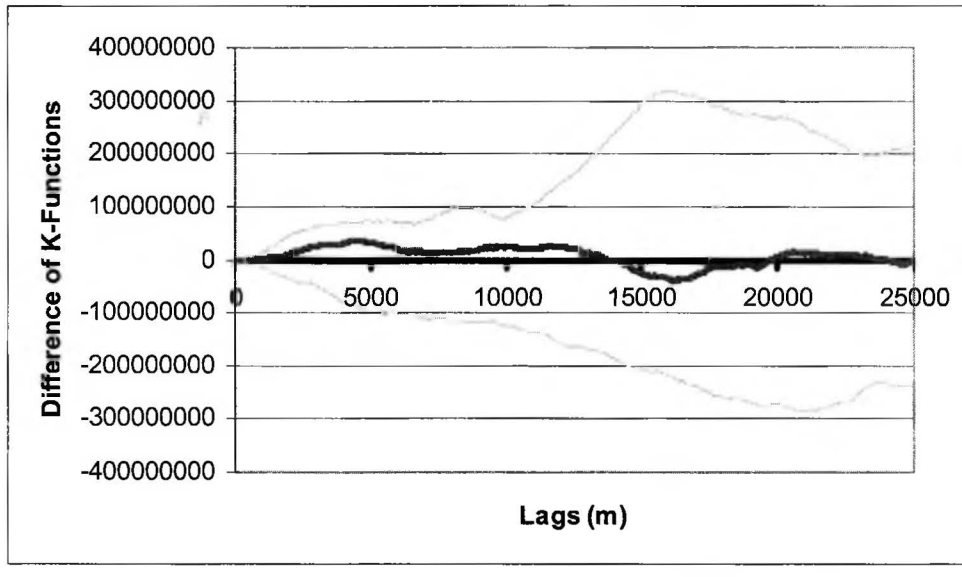


Figure A4.2. Difference of K-Functions for Congenital Anomalies, 1988 – 2002 using a representative point pattern for the control population.

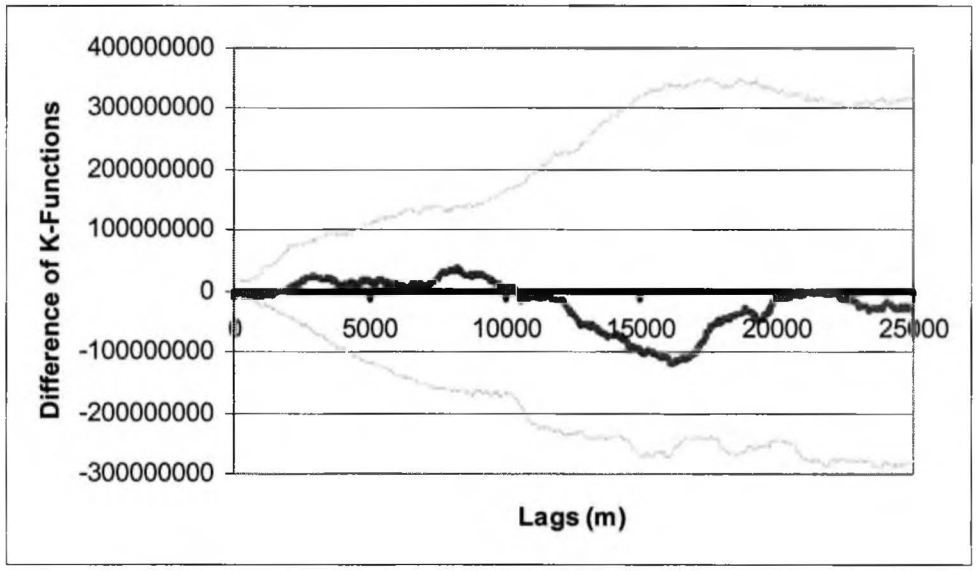


Figure A4.3. Difference of K-Functions for Congenital Anomalies, 1988 – 1992 (n=334 cases)

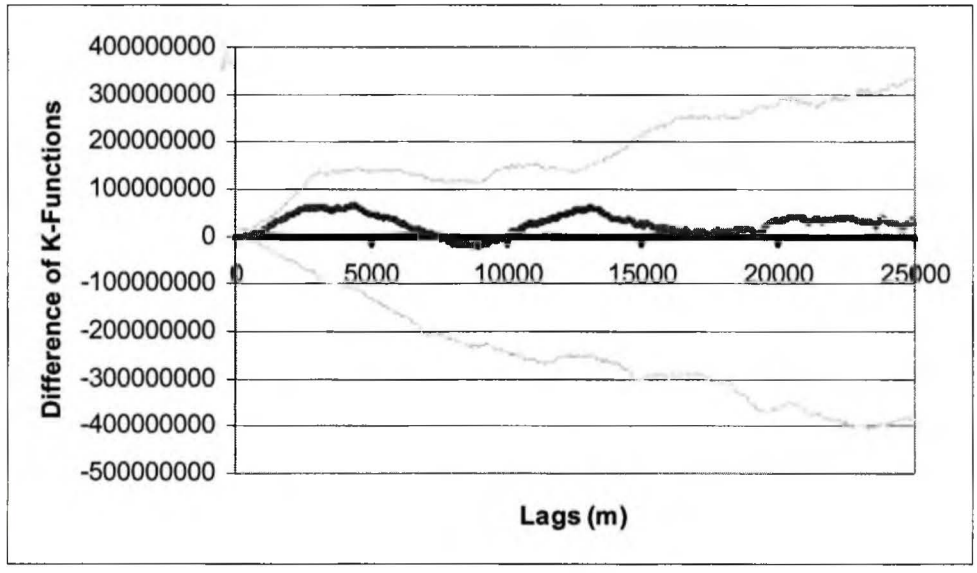


Figure A4.4. Difference of K-Function for Congenital Anomalies, 1993 – 1997. (n=304 cases)

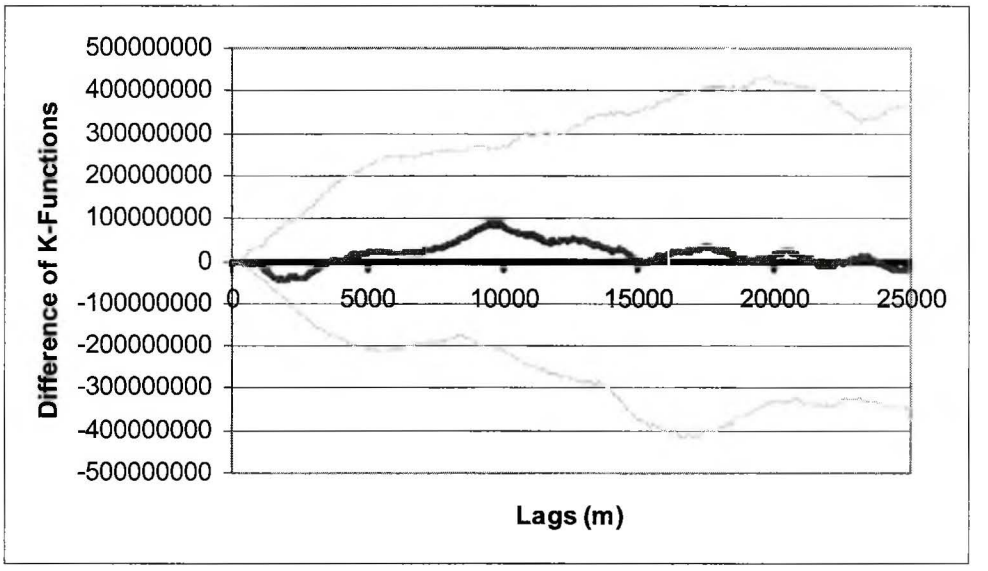


Figure A4.5. Difference of K-Function for Congenital Anomalies, 1998 – 2002. (n = 197 cases)

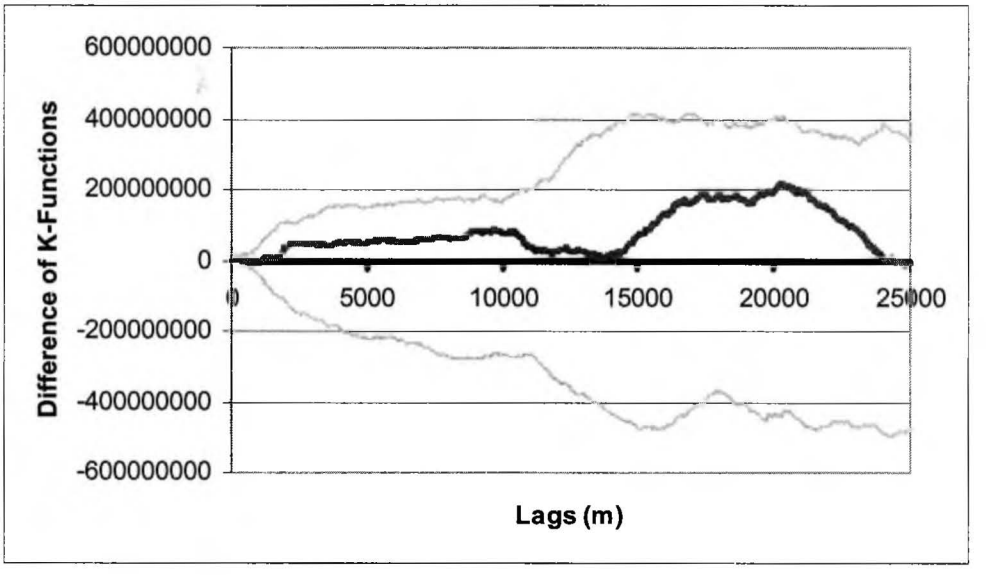


Figure A4.6. Difference of K-Function for LBW, 1988 – 2002.

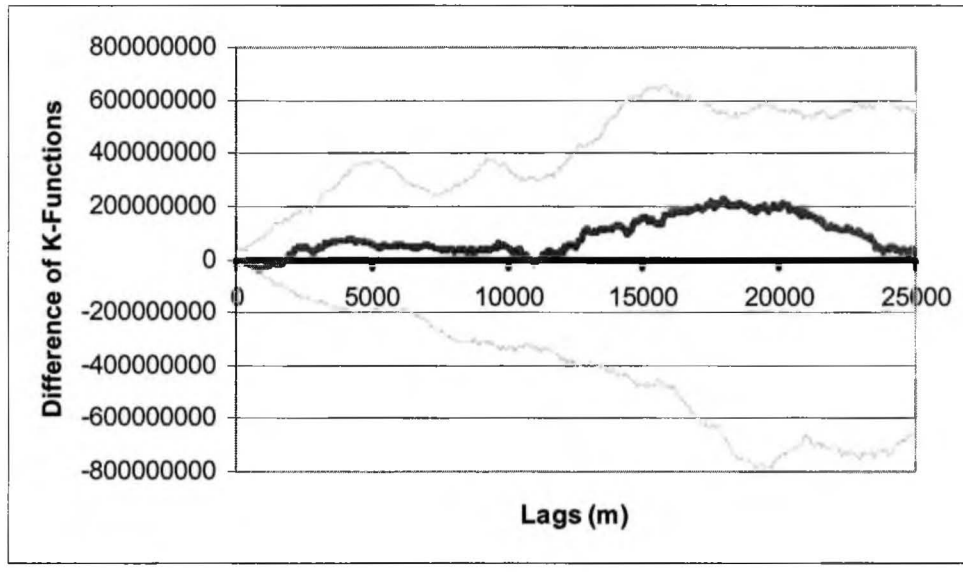


Figure A4.7. Difference of K-Functions for LBW, 1988 – 1992. (n = 83 cases)

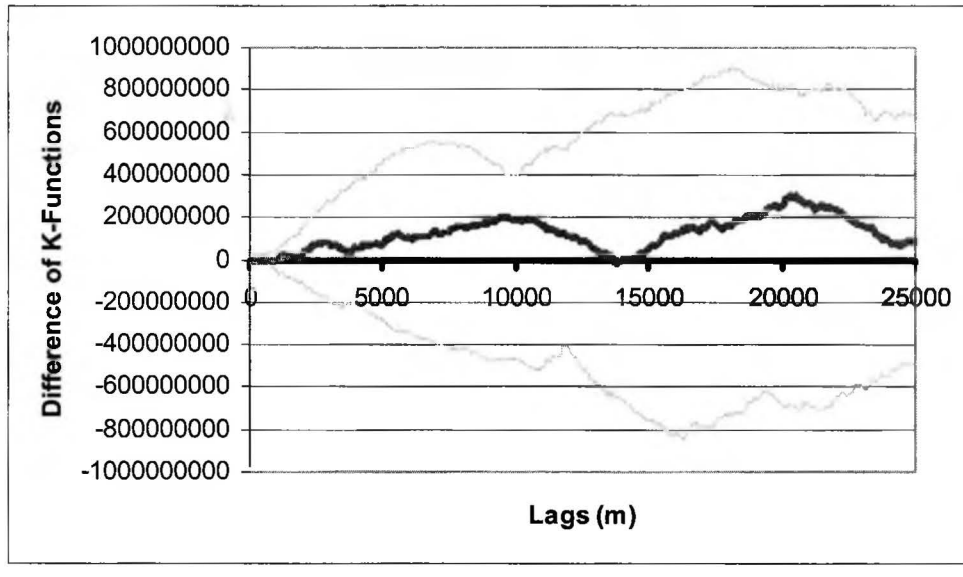


Figure A4.8. Difference of K-Functions for LBW, 1993 – 1997 (n = 67 cases)

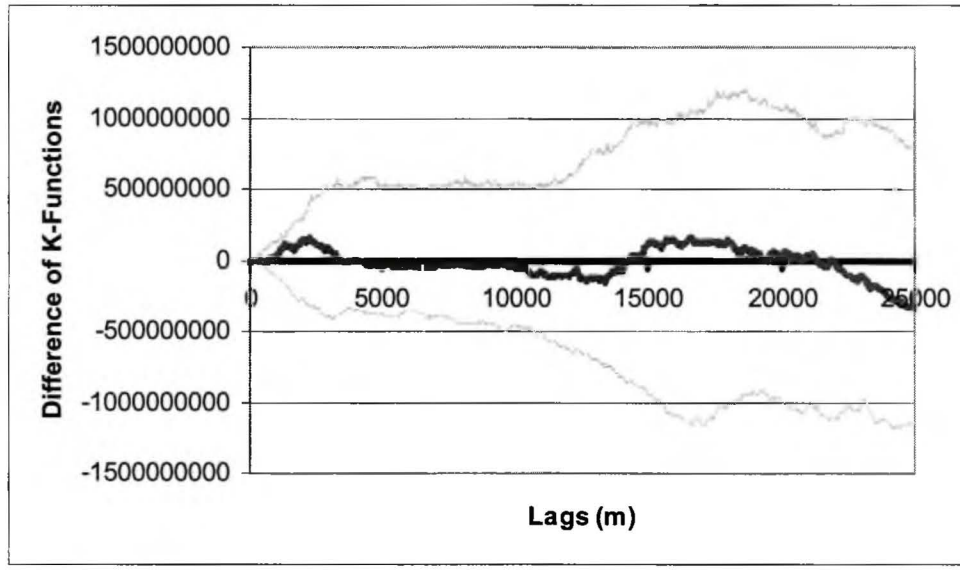
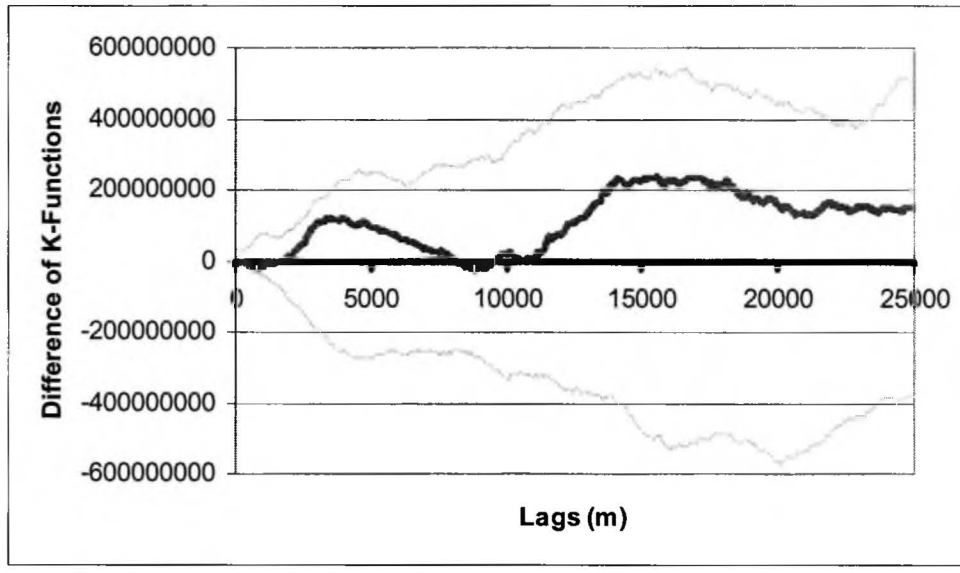


Figure A4.9. Difference of K-Functions for LBW, 1998 – 2002. (n = 35 cases)



A4.10. Difference of K-Functions for Stillbirths, 1988 – 2002.

APPENDIX V

Definitions of Variables Used in Categorical Data Analysis

Table AV.1 List of Variables and the Definitions used in Categorical Data Analysis

Variable Name	Definition
Outcome Variables	
Outcome	Multinomial dependent variable (1 = Preterm, 2 = Congenital Anomaly, 3 = LBW, 4 = Stillbirth, 5 = Normal)
Preterm	Binary dependent variable (0 = Normal, 1 = Preterm)
Anomaly	Binary dependent variable (0 = Normal, 1 = Anomaly)
LBW	Binary dependent variable (0 = Normal, 1 = LBW)
Stillbirth	Binary dependent variable (0 = Normal, 1 = Stillbirth)
Maternal Characteristics	
DMmarry	Binary variable derived from DLMRSTAT (0=Single, Widowed, Divorced, Separated, 1 = Married, Common-Law)
DMmatage	Maternal Age at time of admission (continuous variable)
DMYoung	Maternal Age of 18 years or younger at time of delivery (0 = 19 and Over, 1 = 18 and younger)
DMOld	Maternal Age of 35 years and older at time of delivery (0 = 34 years and younger, 1 = 35 years and over)
DMdeprive	Deprivation index similar to Townsend Index, derived from 1996 Census and converted into a categorical variable based on quintiles (1 = least deprived, 5 = most deprived)
Delivsmk	Derived from the number of cigarettes smoked per day at first prenatal visit, and/or at time of delivery (0 = no, 1 = yes). Initial variable coded 0 for non-smokers, 75 if the patient smoked > 75 cigarettes per day and 88 if patient known to be a smoker, but the number of cigarettes per day was unknown. Variable was recoded to 1 if they were initially coded with a 75 or 88.
CMdrug	Maternal prescription drug use during pregnancy and/or Environmental Exposure (0 = no, 1 = yes). Includes Lithium, maternal exposure to noxious fumes, anti-hypertensives, anti-depressives, anti-epileptics, anti-coagulation therapy, chronic narcotic use
CMabuse	Maternal chemical abuse (0 = no, 1 = yes). Includes alcohol, prescription medication, and narcotic abuse
Pregnancy History Characteristics	
DLpara	Parity (continuous variable) – number of pregnancies excluding the present pregnancy, which resulted in one or more infants weighing 500 grams or more at birth (regardless of whether such infants were stillborn, died after birth or lived
Dlpara04	Binomial variable derived from Dlpara = 0 or > 4 (0 = no, 1 = yes), primiparity or grand multiparity

DLaborts	Binomial variable, any previous pregnancy resulting in a fetus weighing less than 500 grams (non-viable fetus), or when weight not known, less than 20 weeks gestation, regardless of whether the fetus was born alive (0 = no, 1 = yes)
DLprvftd	Binomial variable, any previous pregnancy resulting in a stillbirth (0 = no, 1 = yes), specifically recorded as weighing 500 grams or more and/or equal to or greater than 20 weeks gestation, or when documented as a fetal death by the physician
DLprvlbw	Binomial variable, any previous pregnancy resulting in a LBW (0 = no, 1 = yes)
Dlprev_neg	Binomial variable, any previous pregnancy resulting in an adverse outcome not previously coded (i.e. preterm or congenital anomaly) (0 = no, 1 = yes)
Mpcondition	Binary variable derived by adding all previous maternal conditions (0 = normal, 1 = hypertension (high blood pressure, toxemia, pre-eclampsia or hypertension), anemia, gestational diabetes, eclampsia (convulsions or eclampsia as stated on chart excluding epilepsy)

Present Pregancy Maternal Diagnosis

ADMcomp	Binary variable derived by adding all reasons for admission to hospital. Reasons for admission include: diabetes, intrauterine death, eclampsia (one or more convulsions not attributable to other cerebral conditions such as epilepsy or cerebral hemorrhage in a patient with hypertension), hyperemesis (vomiting which required admission to hospital), suspected fetal anomaly, liver disease, any other complication (0 = no complications, 1 = any of the aforementioned complications)
Mchtd	Binary variable for chronic hypertension (0 = no, 1 = yes). History of hypertensive disease when not pregnant prior to current pregnancy or prior to 20 weeks of current pregnancy; not due to trophoblastic diease—or as stated.
Mpiht	Binary variable for pregnancy induced hypertension (0 = no, 1 = yes)
Manem	Binary variable for anemia (0 = no, 1 = yes). Coded if condition is or was present during current pregnancy. Antepartum HgB <10 gm%)
Mccondition	Binary variable derived by adding the remaining maternal conditions (0 = no, 1 = eclampsia, hyperemesis, fetal growth concerns (suspected or known IUGR or fetal malnutrition), jaundice (fatty liver of pregnancy, serum hepatitis carrier, jaundice of pregnancy/cholestatic liver disease of pregnancy), endocrine (disorder of adrenal gland, disorder of ovary, Hashimoto's Thyroiditis, hyperthyroidism with Goiter, hyperthyroidism with

	Thyroid nodule, hyperthyroidism with Goiter, nodular, hyperthyroidism without Goiter, hypothyroidism, hyperparathyroidism, disorder of Hypothalamus, Disorder of Pituitary gland)

Neonatal Measures

Placenta	Binary variable derived by adding of all placental and umbilical cord anomalies (0 = no, 1 = yes). Includes velamentous insertion of cord, marginal insertion of cord, circumvallate placenta, amnionodosum, succenturiate lobe, chorioamnionitis, funisitis, single umbilical artery
Iwast	Binary variable for Fetal Clinical Soft Tissue Wasting (0 = no, 1 = yes)
Iikdm	Binary variable for Infant of Diabetic Mother (0 = no, 1 = yes)

Exposure Variables Derived in GIS

Within1Km	Binary variable indicating whether observation is within 1 Km from the Coke Ovens site (roughly the geographic center of all point sources of pollution in the area) (0 = no, 1 = yes)
Within2Km	Binary variable indicating whether observation is between 1 and 2 Km away from the Coke Ovens site (0 = no, 1 = yes)
Within3Km	Binary variable indicating whether observation is between 2 and 3 Km away from the Coke Ovens site (0 = no, 1 =yes)
Total3Km	Binary variable indicating whether observation is less than 3 km away from the Coke Ovens site (obtained by summing the previous 3 measures) (0 = no, 1 = yes)
ICB	Categorical variable indicating place of residence (1 = Sydney, 2 = ICB excluding Sydney, 3 = Rest of CBRM)
Dist.to.Coke	Straight line distance of each observation to the Coke Ovens Site (continuous variable)

APPENDIX VI

Results of Categorical Data Analysis

Table A6.1. Odds ratios derived from contingency analyses of adverse reproductive outcomes and proximity to the Coke Ovens site using primiparous cases only (n = 6009)

	DISTANCE BASED MEASURES				COMMUNITY BASED MEASURES		
OUTCOME	**0 - 1 Km** (n=142)	**1 - 2 Km** (n=680)	**2 - 3 Km** (n=498)	**0 - 3 Km** (n=1320)	**Sydney** (n = 1465)	**ICB minus Sydney** (n = 2420)	**Rest of CBRM** (n = 2124)
Preterm Birth							
% Preterm	11.3	10.9	10.2	10.7	10.30	10.20	9.00
p-value*	0.55	0.31	0.73	0.22	0.46	0.44	0.14
Odds Ratio (95% CI)	1.17(0.69 - 1.99)	1.14 (0.88 - 1.48)	1.05 (0.78 - 1.43)	1.13 (0.93 - 1.39)	1.08 (0.89 - 1.31)	1.07 (0.90 - 1.27)	0.87 (0.73 - 1.05)
Number of Cases	16	74	51	141	151	246	192
Congenital Anomaly							
% Anomaly	6.3	4.1	4.8	4.6	4.9	4.8	4.8
p-value*	0.40	0.36	0.99	0.69	0.86	0.92	0.95
Odds Ratio (95% CI)	1.35 (0.68 - 2.67)	0.83 (0.56 - 1.24)	1.00 (0.65 - 1.53)	0.94 (0.71 - 1.26)	1.03 (0.78 - 1.35)	0.99 (0.78 - 1.26)	0.99 (0.78 - 1.27)
Number of Cases	9	28	24	61	72	116	102
Low Birthweight							
% < 2500g	2.1	1.3	1.4	1.4	1.3	1.2	1.5
p-value*	0.39	0.95	0.83	0.61	0.99	0.43	0.41
Odds Ratio (95% CI)	1.23 (0.50 - 5.35)	0.99 (0.60 - 2.06)	1.09 (0.50 - 2.39)	1.15 (0.68 - 1.93)	1.00 (0.59 - 1.68)	0.83 (0.52 - 1.32)	1.21 (0.77 - 1.91)
Number of Cases	3	9	7	19	19	28	31
Stillbirth							
% stillborn	0.7	0.7	0.4	0.6	0.7	0.7	0.5
p-value*	0.87	0.63	0.55	0.97	0.63	0.61	0.34
Odds Ratio (95% CI)	1.18 (0.16 - 8.69)	1.27 (0.49 - 3.278)	0.65 (0.16 - 2.71)	1.02 (0.46 - 2.23)	1.19 (0.58 - 2.48)	1.19 (0.61 - 2.30)	0.70 (0.34 - 1.46)
Number of Cases	1	5	2	8	10	16	10

* p-value determined from a Chi-squared test

Table A6.2. Likelihood Ratio tests for all variables included in the full MNL model.

Likelihood Ratio Tests

Effect	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	4607.914	.000	0	.
Dmmarry	4616.572	8.658	4	.070
DMYoung	4611.461	3.547	4	.471
Dlpara04	4664.851	56.937	4	.000
Dlprvftd	4620.543	12.629	4	.013
Dlprvlbw	4649.986	42.072	4	.000
DLPrev_neg	4667.405	59.491	4	.000
ADMCOMP2	4612.866	4.952	4	.292
Mchtd	4622.259	14.345	4	.006
Mpiht	4633.869	25.955	4	.000
Manem	4633.459	25.545	4	.000
MCcondition	4681.017	73.103	4	.000
Iwast	4832.248	224.334	4	.000
Iikdm	4676.445	68.531	4	.000
Delivsmk	4654.950	47.036	4	.000
CMdrug	4637.939	30.025	4	.000
TOTAL3KM	4620.532	12.618	4	.013
ICB	4621.146	13.232	8	.104

Table A6.3. Classification Table of Outcomes from full MNL model.

Classification

Observed	Predicted					
	1	2	3	4	5	Percent Correct
1	36	0	9	0	1360	2.6%
2	13	0	6	0	743	.0%
3	6	0	2	0	154	1.2%
4	2	0	0	0	78	.0%
5	23	0	6	0	13362	99.8%
Overall %	.5%	.0%	.1%	.0%	99.3%	84.8%

Table A6.4. Likelihood Ratio tests for all variables included in the primiparity MNL model

Likelihood Ratio Tests

Effect	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	870.164	.000	0	.
dmold	874.271	4.107	4	.392
delivsmk	883.667	13.503	4	.009
cmdrug	885.261	15.098	4	.005
mchtd	876.753	6.589	4	.159
mpiht	883.286	13.122	4	.011
iwast	1011.250	141.087	4	.000
iikdm	884.045	13.882	4	.008
NewDeprive	887.692	17.528	16	.352

Table A6.5. Classification table of outcomes from the primiparity MNL model.

Classification

Observed	Predicted					
	1	2	3	4	5	Percent Correct
1	9	0	0	0	580	1.5%
2	3	0	0	0	287	.0%
3	1	0	1	0	76	1.3%
4	0	0	0	0	36	.0%
5	5	0	0	0	5011	99.9%
Overall Percentage	.3%	.0%	.0%	.0%	99.7%	83.6%